

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:18:30 ; Search time 84.6574 Seconds
(without alignments)
1632.058 Million cell updates/sec

Title: US-09-455-978B-2

Perfect score: 2394

Sequence: 1 MSNDNDTLTVADVRNGIDGH.....ATDQQRVVEVRETVGKLS 489

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 29Jan04.*

1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2394	100.0	489	AAE04641	Aae04641 Halobacte
2	412.5	17.2	432	AAE04642	Aae04642 Bacillus
3	399.5	16.7	535	ABU38469	Abu38469 Protein e
4	387.5	16.2	739	AAE04643	Aab96493 Putative
5	386.5	16.1	682	ABU38250	Abu38250 Protein e
6	380	15.9	435	AAE04648	Aab96483 Putative
7	375	15.7	501	AAE04670	Abu96709 Putative
8	374.5	15.6	680	ABU41746	Abu41746 Protein e
9	373.5	15.6	709	ADA33885	Ada33885 Acinetoba
10	371.5	15.5	686	ABU40238	Abu40238 Protein e
11	361	15.1	644	ABU49667	Abu49667 Protein e
12	359.5	15.0	891	AAE26596	Aab26596 Synchocy
13	356.5	14.9	374	AAE04638	Aab96382 Putative
14	356.5	14.9	539	ABU23101	Abu23101 Protein e
15	348	14.5	626	ABU49142	Abu49142 Protein e
16	346.5	14.5	565	ABM68096	Abm68096 Photorhab
17	339.5	14.2	601	ABE49909	Abu49909 Listeria
18	338	14.1	686	ABU24381	Abu24381 Protein e
19	326	13.6	545	ABM68095	Abm68095 Photorhab
20	326	13.6	641	ABU49354	Abu49354 Protein e
21	315	13.2	643	ABU49678	Abu49678 Protein e
22	314.5	13.1	501	ABU41640	Abu41640 Protein e
23	313.5	13.1	541	ABU49123	Abu49123 Protein e
24	308	12.9	845	ABU48593	Abu48593 Protein e
25	298	12.4	511	AAE14148	Abu14148 Bordetell

ALIGNMENTS

RESULT 1
AAE04641

ID AAE04641 standard; protein; 489 AA.

XX AAE04641;

XX

DT 11-SEP-2003 (revised)

DT 04-SEP-2001 (first entry)

XX

DE Halobacterium salinarum HemAT-Hs protein.

XX

KW Haem binding protein; HemAT-Hs; HemAT-Bs; gaseous ligand sensor;
KW oxygen storage; artificial photosynthesis; signalling function;
KW alpha-haemoglobin; myoglobin; therapy.

XX

OS Halobacterium salinarum.

XX

PN WO200140475-A2.

XX

PD 07-JUN-2001.

XX

PF 05-DEC-2000; 2000WO-US033048.

XX

PR 06-DEC-1999; 99US-00455978.

XX

XX (UYHA-) UNIV HAWAII.

XX

PI Alam M, Larsen R;

XX

DR WPI; 2001-374832/39.

XX

NR N-PSDB; AAD08991.

XX

PT Novel isolated bacterial heme binding protein, HemAT-Hs or HemAT-Bs which reversibly binds oxygen with low affinity, useful for controlled storage of oxygen and for sensing gaseous ligands such as oxygen.
PS Claim 6; Page 10; 94pp; English.
CC The present invention relates to isolated archeal and bacterial haem binding protein, HemAT-Hs or HemAT-Bs which reversibly binds oxygen with low affinity. Haem binding protein is useful for controlled storage of oxygen by allowing haem binding protein to bind and store oxygen, and triggering the release of oxygen from haem binding protein by activating the signalling domain. Haem binding protein is useful for sensing gaseous ligands such as oxygen, NO, CO, or CN. Blood substitute comprising haem binding protein is useful for treating a patient suffering from low blood levels by administering and regulating the oxygen binding of the haem-binding protein by modifying the signalling domain. Haem binding protein

26 297.5 12.4 1137 4 ABG24223 Novel hnm
27 289 12.1 565 2 AAW98798 H. pylori
28 289 12.1 565 2 AAW71554 Helicobac
29 289 12.1 565 4 AAB46327 H. pylori
30 284 11.9 630 2 AAW98797 H. pylori
31 284 11.9 630 2 AAW71553 Helicobac
32 284 11.9 675 4 AAB46345 H. pylori
33 280 11.7 675 2 AAY10961 H. pylori
34 277.5 11.6 289 5 ABUS1211 Helicobac
35 276 11.5 293 2 AAW20287 H. pylori
36 273.5 11.4 664 2 AAW20769 H. pylori
37 269 11.2 723 2 AAY19897 B. burgdo
38 269 11.2 753 2 AAY19896 B. burgdo
39 269 11.2 753 6 ABU13306 Protein e
40 266 11.1 606 2 AAY19800 B. burgdo
41 266 11.1 633 2 AAY19799 B. burgdo
42 257 10.7 431 3 AAB52596 Helicobac
43 257 10.7 431 5 ABU50773 Helicobac
44 254 10.6 433 2 AAW05196 Helicobac
45 253 10.6 300 5 ABB49349 Listeria

CC is useful for haem-based catalysis, for artificial photosynthesis and for
 CC identifying potential signalling functions of mutated alpha-haemoglobin
 CC and myoglobin causing several diseases. The present sequence is
 CC Halobacterium salinarium HemAT-Hs protein which is salt tolerant.
 CC (Updated on 11-SEP-2003 to standardise OS field)
 XX

SQ Sequence 489 AA;

Query Match 100.0%; Score 2394; DB 4; Length 489;
 Best Local Similarity 100.0%; Pred. No. 6.3e-165;
 Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNDNDTLVTADVRNGIDGHALADRIGLDEARIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
 DB 1 MSNDNDTLVTADVRNGIDGHALADRIGLDEARIAWRLSFTGIDDDTMAALAAEQPLFEAT 60

QY 61 ADALVTDFYDHLSEYRTQDLFANSTKTKVEQLKETOABYLLGLGRGEYDTEYAAQRARIG 120
 DB 61 ADALVTDFYDHLSEYRTQDLFANSTKTKVEQLKETOABYLLGLGRGEYDTEYAAQRARIG 120

QY 121 KIHVDVLGPGDVLGAYTRYTYTGLLDALDDVADRGHEAAAVDELVARFLPMLKLTFF 180
 DB 121 KIHVDVLGPGDVLGAYTRYTYTGLLDALDDVADRGHEAAAVDELVARFLPMLKLTFF 180

QY 181 DQOIAMDYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSDVAERTDTMRART 240
 DB 181 DQOIAMDYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSDVAERTDTMRART 240

QY 241 DQVDRMADVSRISVSAEVEASTADDVVRTSEDAPALAAQGEAAADDAATMTDID 300
 DB 241 DQVDRMADVSRISVSAEVEASTADDVVRTSEDAPALAAQGEAAADDAATMTDID 300

QY 301 EATDGTAGVQELGRADVESVTGVIDDIABOTNMALNASIEAARAGEAGEFAVVAD 360
 DB 301 EATDGTAGVQELGRADVESVTGVIDDIABOTNMALNASIEAARAGEAGEFAVVAD 360

QY 361 EVKALAEBSREOSTRVEELVEQMAETETVDQLDEVNQRIGEGVERVEEAMETLQETID 420
 DB 361 EVKALAEBSREOSTRVEELVEQMAETETVDQLDEVNQRIGEGVERVEEAMETLQETID 420

QY 421 AVEDAASGMQVSTATDQAVSTERVAEWDGVDVDRAGEIAAALDDIADATDQQVTVVEE 480
 DB 421 AVEDAASGMQVSTATDQAVSTERVAEWDGVDVDRAGEIAAALDDIADATDQQVTVVEE 480

QY 481 VRETGKLS 489
 DB 481 VRETGKLS 489

RESULT 2

AAE04642
 ID AAE04642 standard; protein; 432 AA.

XX AC

XX AAE04642;

XX 04-SEP-2001 (first entry)

XX DE Bacillus subtilis HemAT-Bs protein.

XX Haem binding protein; HemAT-Hs; HemAT-Bs; gaseous ligand sensor;
 KW oxygen storage; artificial photosynthesis; signalling function;
 KW alpha-haemoglobin; myoglobin; therapy.

XX OS Bacillus subtilis.

XX WO200140475-A2.

XX 07-JUN-2001.

XX 05-DEC-2000; 2000WO-US033048.

XX 06-DEC-1999; 99US-00455978.

XX

PA (UYHA-) UNIV HAWAII.

XX Alam M, Larsen R;

XX WPI; 2001-374832/39.

XX N-PSDB; AAD08992.

XX Novel isolated bacterial heme binding protein, HemAT-Hs or HemAT-Bs which

XX reversibly binds oxygen with low affinity, useful for controlled storage

XX of oxygen and for sensing gaseous ligands such as oxygen.

XX Claim 8; Page 11; 94pp; English.

XX The present invention relates to isolated archeal and bacterial haem

XX binding protein, HemAT-Hs or HemAT-Bs which reversibly binds oxygen with

XX low affinity. Haem binding protein is useful for controlled storage of

XX oxygen by allowing haem binding protein to bind and store oxygen, and

XX triggering the release of oxygen from haem binding protein by activating

XX the signalling domain. Haem binding protein is useful for sensing gaseous

XX ligands such as oxygen, NO, CO, or CN. Blood substitute comprising haem

XX binding protein is useful for treating a patient suffering from low blood

XX levels by administering and regulating the oxygen binding of the haem-

XX binding protein by modifying the signalling domain. Haem binding protein

XX is useful for haem-based catalysis, for artificial photosynthesis and for

XX identifying potential signalling functions of mutated alpha-haemoglobin

XX and myoglobin causing several diseases. The present sequence is Bacillus

XX subtilis HemAT-Bs protein

XX SQ Sequence 432 AA;

Query Match 17.2%; Score 412.5; DB 4; Length 432;

Best Local Similarity 26.2%; Pred. No. 1.4e-21;

Matches 114; Conservative 85; Mismatches 187; Indels 49; Gaps 8;

QY 31 AEIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFYDHLSEYRTQDLFANSTKTV 90

DB 32 ADVKQLKMWRLGDAEYLVLEQLQIENIVNIDAFYKNLDHSSLMDDII-NDHSSVD 90

QY 91 QLKETQAEYLLGLGRGEYDTEYAAQRAEIKTHDVLGLGPDVYLGAITYTYTGLLDALAD 150

DB 91 RLKQTLKRIHQEMFAGVIDDFIEKRNRIASIHRLIGLLPKWYMGAFQELLMSMDIY-- 148

QY 151 DWADRGEEAAAADVELVARFLPMLKLTDFQQLAMDTYIDSYAQRLHDEIDSRQELANA 210

DB 149 -----EASITNQQELLKATTKILNLQELVLEAFQSEYNQ-TRDEQEEKNLH- 199

QY 211 VATHVEAPLSSLEATSDVAERTDTMRARTDQVDRMADVSRISVSAEVEASTADD 270

DB 200 -----QKIQTSGSIA-----NLFSETSRVQELVDKSEGISQASKAGTTSSTVEE 246

QY 271 VRTSDEAEALAAQGEAAADDAATMTDIDEATDGTAGVQELGERAADVSVTVGIDDI 330

DB 247 -KSIGGKKELVQKQ-----MNKIDTSLVQIEKEMVKLDEIAQQIEKIFGIVTGI 296

QY 331 AEQTNMLNALNASIEAARAGEAGEFAVVADDEVKALAEBSREOSTRVEELVEQMAETET 390

DB 297 AEQTNLLSLNASIESARAGEHKGKGFVAVNEVRKLSEOTKKTSTVSELVNTNTQINIV 356

QY 391 VDQLDEVNQRIGEGVERVEEAMETLQETIDDAVEDA-----ASGMQVSVSTAT 436

DB 357 SKHINDVNLVSESEKEMKTQINRLFDTEIVHSMKISKEQSGKIDVDLQAFGLGQVSRV 416

QY 437 DEQAVSTEEVAEMVD 451

DB 417 SHVAASVDSLILTE 431

RESULT 3

ABU38469

ID ABU38469 standard; protein; 535 AA.

XX

XX ABU38469;

XX

19-JUN-2003 (first entry)
 Protein encoded by Prokaryotic essential gene #23996.
 Antisense; prokaryotic essential gene; cell proliferation; drug design.
 Pseudomonas aeruginosa.
 WO200277183-A2.
 03-OCT-2002.
 21-MAR-2002; 2002WO-US009107.
 21-MAR-2001; 2001US-00815242.
 06-SEP-2001; 2001US-00948993.
 25-OCT-2001; 2001US-0342923P.
 08-FEB-2002; 2002US-00072851.
 06-MAR-2002; 2002US-0362699P.
 (ELIT-) ELITRA PHARM INC.
 Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,
 Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 WPI; 2003-029926/02.
 N-PSDB; ACA42339.
 New antisense nucleic acids, useful for identifying proteins or screening
 for homologous nucleic acids required for cellular proliferation to
 isolate candidate molecules for rational drug discovery programs.
 Claim 25; SEQ ID NO 66393; 1766pp; English.
 The invention relates to an isolated nucleic acid comprising any one of
 the 6213 antisense sequences given in the specification where expression
 of the nucleic acid inhibits proliferation of a cell. Also included are:
 (1) a vector comprising a promoter operably linked to the nucleic acid
 encoding a polypeptide whose expression is inhibited by the antisense
 nucleic acid; (2) a host cell containing the vector; (3) an isolated
 polypeptide or its fragment whose expression is inhibited by the
 antisense nucleic acid; (4) an antibody capable of specifically binding
 the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 proliferation or the activity of a gene in an operon required for
 proliferation; (7) identifying a compound that influences the activity of
 the gene product or that has an activity against a biological pathway
 required for proliferation, or that inhibits cellular proliferation; (8)
 identifying a gene required for cellular proliferation or the biological
 pathway in which a proliferation-required gene or its gene product lies
 or a gene on which the test compound that inhibits proliferation of an
 organism acts; (9) manufacturing an antibiotic; (10) profiling a
 compound's activity; (11) a culture comprising strains in which the gene
 product is overexpressed or underexpressed; (12) determining the extent
 to which each of the strains is present in a culture or collection of
 strains; or (13) identifying the target of a compound that inhibits the
 proliferation of an organism. The antisense nucleic acids are useful for
 identifying proteins or screening for homologous nucleic acids required
 for cellular proliferation to isolate candidate molecules for rational
 drug discovery programs, or for screening homologous nucleic acids
 required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
K. pneumoniae or *P. aeruginosa*. The present sequence is encoded by one of
 the target prokaryotic essential genes. Note: The sequence data for this
 parent did not form part of the printed specification, but was obtained
 in electronic format directly from WIPO at
 ftp.wipo.int/pub/published_pct_sequences
 Sequence 535 AA;
 Query Match 16.7%; Score 399.5; DB 6; Length 535;
 Best Local Similarity 25.2%; Pred. No. 1.6e-20;
 Matches 132; Conservative 111; Mismatches 206; Indels 75; Gaps 16;
 3 NDNDTLTVADVNRGIDGHALADRLGLDEAETAWRLSFTGIDDDTMAALAAEQPLFEATAD 62

Db 37 SENELSVNA-LRNHEGDMMDH-----ALRADVLAAFW-OPDGAARE 78
 Qy 63 ALVTFDYDHLSEYERTQDLFANSTKTVEQ-----LKETQAEYLLGLGRGYDTEYAAQRA 117
 Db 79 QVRQDLQEHSGWFR-----KVVEQNGQLFLNDIAHQALVEL-RPDLEAYIGAAS 127
 Qy 118 RIGK-IHDVLGLGPDV--YLGATRYTYTGLLDALADV-----VADRGEE----- 159
 Db 128 IVGKALLDPVAARAELPFQVQAF-KBLEGRNEALSSLIEKHVEQTNAREDSMRYSAWML 186
 Qy 160 -----AAAAVDELVARFL-----PMLKLTFFDOQIAMDTY-----IDS--YAORLHDEI 201
 Db 187 AGGILVACLVLGQLCRQLLRAVLQPLRLKLVSSARVIAQGNLOEPIGVDSNDEAQLQAL 246
 Qy 202 DSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVASV 261
 Db 247 GEMQENLRMITIIRQSESEELHDTSGISGOTSQSIHVHGASQQAADSATSMAASMEEMITNI 306
 Qy 262 EVASTADDDVPTSEDAAEAQAQGEAAADALATMTDIDEATDGTAGVQEOGGERADVE 321
 Db 307 SQISDHADNARVISAKEELASSGGQVILNVVEGMSRIADVNVQSSTSI TALGQSSDEIH 366
 Qy 322 SVTGVIDDIAEOTNMLNLSIAEARGEAGEGFAVVADEVKALAEESREGSTRVEELVE 381
 Db 367 SIIQVIKIGIAEQTNLLNLAIEAARAGEAGRGFAVVADEVGLAARTTQTQITAMIE 426
 Qy 382 QMQAETETVDQDEVNORIGEGVERVEEAMETLQETDAVEDAASGMQEVSTATDEQAV 441
 Db 427 RIRASTGQAINMEAGVSRVNEGVSPARQAGVSNELIDGTRHAASVVDEISQITIREQSR 486
 Qy 442 STEVAEMVDGVDVRAGEIAAALDDIADATDQOVRTVEEVRETV 485
 Db 487 ASDEIAQRVELIAQRSQONTQAMHEMA-AT---ARRLNEVAATM 526
 RESULT 4
 AAB96493
 ID AAB96493 standard; protein; 739 AA.
 AC AAB96493;
 XX 29-OCT-2001 (first entry)
 DE Putative sensory transduction histidine kinase and response regulator #3.
 XX Hyperthermophilic archaeon; hyperthermophilic protein.
 XX Pyrococcus abyssi.
 XX FR2792651-A1.
 XX 27-OCT-2000.
 XX 21-APR-1999; 99FR-00005034.
 XX 21-APR-1999; 99FR-00005034.
 XX (CNRS) CNRS CENT NAT RECH SCI.
 XX (IFRE-) IFREMER INST FR RECH EXPL MER.
 PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
 PI Querellou J, Weissenbach J, Saurin W, Heilig R;
 XX WPI; 2001-126236/14.
 XX New nucleotide sequences isolated from *Pyrococcus abyssi* encode proteins
 useful in industry.
 XX Claim 7; Page 1203-1205; 1657pp; French.
 CC The present invention relates to the genomic sequence of *Pyrococcus*
abyssi (see AAF86431 and AAH41223-7) and *P. abyssi* proteins. *P. abyssi* is

Variable	Mean	Standard Deviation	Minimum	Maximum
Age	35.2	12.5	22	65
Gender	0.52	0.50	0	1
Marital Status	0.68	0.48	0	1
Education	12.8	2.1	9	16
Income	45000	15000	20000	80000
Health	0.75	0.43	0	1
Smoking	0.25	0.43	0	1
Alcohol	0.15	0.36	0	1
Exercise	0.35	0.48	0	1
Stress	0.65	0.48	0	1
Depression	0.12	0.33	0	1
Loneliness	0.28	0.45	0	1
Social Support	0.72	0.45	0	1
Life Satisfaction	0.58	0.49	0	1
Resilience	0.62	0.48	0	1
Optimism	0.68	0.47	0	1
Gratitude	0.75	0.43	0	1
Forgiveness	0.65	0.48	0	1
Self-Compassion	0.55	0.49	0	1
Emotional Regulation	0.62	0.48	0	1
Mindfulness	0.45	0.50	0	1
Transcendental Meditation	0.35	0.48	0	1
Vipassana	0.25	0.43	0	1
Zen	0.15	0.36	0	1
Therapy	0.12	0.33	0	1
Support Group	0.08	0.27	0	1
Religious Practice	0.22	0.41	0	1
Community Involvement	0.38	0.49	0	1
Volunteering	0.28	0.45	0	1
Charitable Giving	0.18	0.38	0	1
Philanthropy	0.12	0.33	0	1
Environmental Activism	0.08	0.27	0	1
Animal Welfare	0.05	0.22	0	1
Human Rights	0.03	0.18	0	1
Peace Activism	0.02	0.14	0	1
Anti-Corruption	0.01	0.10	0	1
Transparency	0.01	0.10	0	1
Accountability	0.01	0.10	0	1
Integrity	0.01	0.10	0	1
Honesty	0.01	0.10	0	1
Trust	0.01	0.10	0	1
Respect	0.01	0.10	0	1
Equality	0.01	0.10	0	1
Justice	0.01	0.10	0	1
Freedom	0.01	0.10	0	1
Democracy	0.01	0.10	0	1
Rule of Law	0.01	0.10	0	1
Corruption	0.01	0.10	0	1
Abuse of Power	0.01	0.10	0	1
Discrimination	0.01	0.10	0	1
Oppression	0.01	0.10	0	1
Exploitation	0.01	0.10	0	1
Violence	0.01	0.10	0	1
War	0.01	0.10	0	1
Conflict	0.01	0.10	0	1
Peace	0.01	0.10	0	1
Stability	0.01	0.10	0	1
Security	0.01	0.10	0	1
Prosperity	0.01	0.10	0	1
Wealth	0.01	0.10	0	1
Poverty	0.01	0.10	0	1
Unemployment	0.01	0.10	0	1
Unemployment Rate	0.01	0.10	0	1
Unemployment Duration	0.01	0.10	0	1
Unemployment Benefits	0.01	0.10	0	1
Unemployment Insurance	0.01	0.10	0	1
Unemployment Compensation	0.01	0.10	0	1
Unemployment Allowance	0.01	0.10	0	1
Unemployment Support	0.01	0.10	0	1
Unemployment Assistance	0.01	0.10	0	1
Unemployment Relief	0.01	0.10	0	1
Unemployment Help	0.01	0.10	0	1
Unemployment Aid	0.01	0.10	0	1
Unemployment Grant	0.01	0.10	0	1
Unemployment Stipend	0.01	0.10	0	1
Unemployment Allowance	0.01	0.10	0	1
Unemployment Support	0.01	0.10	0	1
Unemployment Assistance	0.01	0.10	0	1
Unemployment Relief	0.01	0.10	0	1
Unemployment Help	0.01	0.10	0	1
Unemployment Aid	0.01	0.10	0	1
Unemployment Grant	0.01			

QY 228 DVARTDTR-----ARTDDQVDMADVREISSVSASVEVA----- 265
 DB 198 NLRLMKTVRDLALTLEGRANDLTRISSEISEAINQVABATQVSVQROQENITEIME 247
 QY 266 ---STADVRRTSDEAE-----ALAOQGEAAADALATMTDIDEATDGTAGVEQ 312
 DB 248 GWNITADVTQRTVDAMEEFGVNVNLSIAREGDKGKEAISQVEDIQDMKVIRQAOVE 307
 QY 313 LGERAADVESVTGVDDIAEQTNMLNLSIAEAPAGEGFAVVADEVKALAEBSREQ 372
 DB 308 VAEMSKNVGDIINALADIAEQTNMLNLSIAEAPAGEGFAVVADEVKALAEBSREQ 367
 QY 373 STRVEELVEQMAETEEFTVDQDEVNQIRGCVREVEAMETLQBITDAVEDAAGMVEV 432
 DB 368 AEKIRGILNEIQEKVAKVEETKGVKVVDSVDFLKTETVGLMNMIGELDDVESKLQDI 427
 QY 433 STATDEQAVSTEVEAMVDGVDDDRAGETAAALDDIADATDQQRVTVEEVRVTKL 488
 DB 428 KNELANTOEHVENAKALENLAASAQETTASAEVSSASQAQSSMEVEVKRNIIEL 483

RESULT 8

ABU41746
 ID ABU41746 standard; protein; 680 AA.

AC ABU41746;

DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #27273.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Pseudomonas syringae.

OS WO200277183-A2.

PN 03-OCT-2002.

PD 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR N-PSDB; ACA45616.

XX Claim 25; SEQ ID NO 69670; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation and the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 680 AA;

Query Match 15.6%; Score 374.5; DB 6; Length 680;
 Best Local Similarity 24.2%; Pred. No. 1.4e-18;
 Matches 120; Conservative 103; Mismatches 201; Indels 71; Gaps 12;

QY 23 ADRIGLDEAEIA-----WRLSFTGIDD-DTMAALAAEQPLFE---ATADALVTDF 68
 DB 207 ADAFGRDASQFGRVLNGLMGNATLRIITQVEDRARARLARAEIAELFEVFSVSDILRTS 266
 QY 69 YDHLSEYERTODLFANSTKTVEQLKETQAEYLLGLGRGEYDEYAAQARIGKIH-D-VIG 127
 DB 267 PELYQVREASGNIP-NTSQTLDDTSVLANSI-----ENLAKRRTWNTVGGYVIG 315
 QY 128 LGPDVYLGAITYTYTGLLDALADDVADRGEAAAVDELVARPLPMLKLLTFDQQIA-- 185
 DB 316 L-----LALMSIILIGLV-----MVRETNRQLRETAQKSERNQTAIMRLDIEIENLADG 364
 QY 186 -----MDYIIDSYAQRLHDEIDSRQELA---NAVATHVEAPLSSLEATSDQVAERTD 234
 DB 365 DLTVTASVTEDFTGAIDSNYSIDQLRELVTINLTAEQVASAVTETQATAMQLS---- 420
 QY 235 TMRARTDDQVDRMADVREISSVSASVEEASTADDVRRTSDEADALAQOGEAAADALA 294
 DB 421 ---AASEHQALQISAASTAVNDMAASIDQVSAASESSAVERSVAIANKGVNVQNTIH 477
 QY 295 TMTDIDEATDGTAGVEQLGERAADVESVTGVDDIAEQTNMLNLSIAEAPAGEGAGEG 354
 DB 478 GMDNIREQIQDTSKRIKLGESSQEGIDVLSLDDIADQTNILALNAAIQASMGADGRG 537
 QY 355 FAVVADEVKALAEBSREQSTRVEELVEQMQOAEETETVDQLDEVNORIGEGYRVESEMET 414
 DB 538 FAVVADEVQRLAERSSSSATKQIETLVRAIQNDTNEAVISMETQTTSEVVGRGARLAQDAGVA 597
 QY 415 LQETDAVEDAASGMQEVSTATDEQAVSTEVEAEVMDGVDDDRAGEIAAALDDIADATDQ 474
 DB 598 LQETEGSVRLAEILIESITDAHQQAES-----AGQISQTMVTVIQTTSTQ 643
 QY 475 VRTVEEVRVETVGKLS 489
 DB 644 TSGTSATAESIGNLA 658

RESULT 9

ADA33885

ID ADA33885 standard; protein; 709 AA.

XX ADA33885;

AC ADA33885;

XX DT 20-NOV-2003 (first entry)

XX Acinetobacter baumannii protein #1046.

XX Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
KW plant biocontrol agent.
XX Acinetobacter baumannii.
OS US6562958-B1.
FN 13-MAY-2003.
XX
XX 04-JUN-1999; 99US-00328352.
XX
XX 09-JUN-1998; 98US-0088701P.
PR (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Breton G, Bush D;
PI
XX WPI; 2003-576092/54.
DR N-PSDB; ADA29759.
XX
XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX
XX Example; SEQ ID NO 5172; 328pp; English.
XX
XX The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents the amino acid sequence of an A.
CC baumannii protein.
XX
XX Sequence 709 AA;

Query Match 15.6%; Score 373.5; DB 6; Length 709;
Best Local Similarity 28.7%; Pred. No. 1.7e-18;
Matches 100; Conservative 67; Mismatches 141; Indels 41; Gaps 5;
QY 144 LLDAADVADRGEEAAAADVELVARPLMKLLTFDQIAMTYDSDVAQRHLRIDS 203
Db 388 LLEIAD--LADGLRSTATVSE-----DFTGATADSNFADQ 424
QY 204 RQELANAVATHVEAPLSLEATSQDVAERTDTMPARTDDQVDRMADVSRITSSVSASVEE 263
Db 425 LRDL-----VSRIHETSQEVARYTQDTQSIINQLAAESEHQAEIAGASTANNE 473
QY 264 VASTADV-RTSDEAA-----LAQGEAAADALATWTDIDEATDGVTAQVEQLGER 316
Db 474 MAQSIDQVSANASASAEVAQRVSQIASNGAQNVRNRSIEGMTIREQIQTSKRIKLGES 533
QY 317 AADVESVTGVIDDAEQTNNLANASTAEAPAGEGFAVVADEVKALAESEREQSTRV 376
Db 534 SQEIGNVSLINDIADQTNILANAAIQASVAGAGRGFAVVADEVQRLAERSASATKQI 593
QY 377 BELVEQMAETEFTVDQLDEVQNRIGBGVERVEAMETLQBITDAVEDAASGMQEVSTAT 436
Db 594 ETLVKTIQTDNEAVISMEQTTEVRGANLAKDAGIALDEIQKVSGLAKLIASIDAA 653
QY 437 DEQAVSTEEVAMVDGVDRAGETAAALDDIADATDQQVTRVERTV 485
Db 654 KLOSASASHIATTTVVVQETISQTTTATFOTARSVELANMAESRESV 702

RESULT 10

ABU40238

ID ABU40238 standard; protein; 686 AA.

XX

AC ABU40238;
XX
XX 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #25765.
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
KW Pseudomonas putida.
XX
XX WO200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
DR N-PSDB; ACA44108.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 68162; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 686 AA;

Query Match 15.5%; Score 371.5; DB 6; Length 686;
Best Local Similarity 24.7%; Pred. No. 2.3e-18;
Matches 119; Conservative 91; Mismatches 196; Indels 75; Gaps 11;


```

Db      19 GKDYENSSNVTQAIINLYLDVLDGKPEQPLG-----LSKXDEB-- 59
Qy      162 AAVDELVARFLPMLKLLTFDQOQIMDTVIDSYAQRLLHDEIDSRQELANAVATHVEAPLSS 221
Db      60 -----VLKKVAFRLKRGQKKIN-----VKDKIENLKE-----VIENLEEKIG- 97
Qy      222 LEATSQVAERTDMRATDQVDRMADVSRISVSASVEEASTADDVRTSEDAEAL 281
Db      98 -EVLKGLDGE-VNELVSLNNENKIAENVYIQTLSAGIEMNVQAOLSDFALESASM 155
Qy      282 AQQGEAADDALATMTDIDEATDGTAGVQELGERAADVESVTGVIDDIAEQTNMLANA 341
Db      156 AEKGRQISDNVALKVRSISREMSDAVRLAEVSKKINDIVYVVISIAQTNMLANA 215
Qy      342 STEARAGEGEGFAVVADEKALAEBSREOSTRVEELVEQQAETETVDQDEVNORI 401
Db      216 STEARAGEAGRFVAVVAENVRELADRSKGAEEIRNIIEMQ-----ENINRVIQAI 268
Qy      402 GEGVERVEEAMETLOEITDAVED-----AASGMQEVSTATDEQAVSTEVEAEVMDGVD 454
Db      269 QENVRTVEEVAIONLTAADDIARRANETANMVKELSEGIDEQANSVQMLVDKIDSIS 328
Qy      455 DRAGEIAALDDIADATDQQVRTVEVRETGVKL 488
Db      329 KDVSNNLNFATQLTDTISSLEKLNKVEEITKL 362

```

RESULT 14

ABU23101
ID ABU23101 standard; protein; 539 AA.

XX AC ABU23101;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #8628.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Bordetella pertussis.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX DR N-PSDB; ACA26971.

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 51025; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of

XX CC the 6213 antisense sequences given in the specification where expression

XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:

XX CC (1) a vector comprising a promoter operably linked to the nucleic acid

XX CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *X. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 539 AA;

Query Match 14.9%; Score 356.5; DB 6; Length 539;
Best Local Similarity 26.6%; Pred. NO. 2.1e-17;
Matches 126; Conservative 81; Mismatches 153; Indels 113; Gaps 15;
Qy 38 SFTGIDDDTMAAL-----AAEQPLFEATADALVTDYDHLSEYERTQDLFANSTKTVQQL 92
Db 126 SFAASLDEMMAALERNDDAAYLQKNVKAQASAAFAARLGEFSTNLDKLSSETLAHET 185
Qy 93 KETQAEYLLGLGRGEYDTEYAAQRAIRIKIHVILGPDVYLGYTRYTYTGLLDALADV 152
Db 186 RETIMLYV-----YAA-----LLLLIVGVIAGSYLYMTRAVVRPLQR-- 222
Qy 153 VADRGEEAAAADVDELVARFLPMLKLLTFDQOIAMDTYIDSYAQRLLHDEIDSRQELANAVA 212
Db 223 -ASQOFRMAAGD-LSARI-----EGNSRNEIGALIT 252
Qy 213 T--HVEAPLS-----SLEATSQDVAERTTMEARTDDQVDRMADYSRBISSVSA 259
Db 253 ALRHMQESLTRTAAVRRGVDEINVGSREISAGNTDLSRTEEQ-----AASLEETAA 305
Qy 260 SVEEVAST-----ADDVRTSE---DAEALAQCGEAAAADDALATMTDIDEATDGVTAGVEQ 312
Db 306 SMEQLASTVKQNADNARQANQALASASDVESGGSVSEVWSTM-----DGISASSRK 358
Qy 313 LGERAADVESVTGVIDDIAEQTNMLNALNASTEAARAGAGBGFVAVVADDEVKALAEBSREQ 372
Db 359 ISE-----IVSVIDGIAFQTNILNALNAVEAARAGEQKGFVAVVAGEVRSLAQRSAQA 411
Qy 373 STRVEELVEQQAETETVDQLDENVQRIQGVVERVEEAMETLOEITDAVEDAASGMORV 432
Db 412 AKBIKVLIE-----DSVD-----KVGTSQOQVERAGATMQEIVASVKRVTDIMGEI 457
Qy 433 STATDEQAVSTEVEAEVMVDGVDRAGETAAALDDIADATDQQVRTVEEVRETIV 485
Db 458 SAASESQSGIIEQVNRVNSQMDVETQQNALVEEAAAAGSLQEQAQRLEAAV 510

RESULT 15

ABU49142

ID ABU49142 standard; protein; 626 AA.

XX AC ABU49142;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #34669.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Vibrio cholerae.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX EF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX XW Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX DR WPI; 2003-029926/02.
XX DR N-PSDB; ACA53012.
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 25; SEQ ID NO 77066; 1766pp; English.
XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the
XX CC proliferation of an organism. The antisense nucleic acids are useful for
XX CC identifying proteins or screening for homologous nucleic acids required
XX CC for cellular proliferation to isolate candidate molecules for rational
XX CC drug discovery programs, or for screening homologous nucleic acids
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
XX CC the target prokaryotic essential genes. Note: The sequence data for this
XX CC patent did not form part of the printed specification, but was obtained
XX CC in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 626 AA;

Query Match 14.5%; Score 348; DB 6; Length 626;
Best Local Similarity 27.4%; Pseq No. 1e-16;
Matches 92; Conservative 78; Mismatches 134; Indels 32; Gaps 6;
174 MLKLLTFDQOI-----AMDTYIDSYAQRHLHDEIDSRQELANAVATHEAPLSS 221
| | | | | : | | | | | : | | : : |

Db 298 MSPLKTLDSAIKDIAAGGDLTKKLDITNLDKFESELAGFNSTFTEMLGSGIROLKTIASG 357
Qy 222 LEATSDQVAERTDTMRARTDDQVDRMDVSRSEISSVASVEEASTADDVVRTSEDAEAL 281
Db 358 VLDGAEKTAANEAEVSRSLVVEQQL-----QELQLATAMNEMAMTASEVANSQAQVAADA 410
Qy 282 AOCGEAAADDAATMTDIDEATDGVV-----AGVEQLGE--RAAD-VESVTGVVIDDIA 331
Db 411 AKGESA---SLEGSSVVHETTTDAIORLSTRIGSSVEDVKELVRATRIETVLDVINDIA 467
Qy 332 EOTNMLNANASIEAARAGEAGEFAVVADEVKALAEESREOSTRVEELVEQMOAETETV 391
Db 468 DOTNLLALNAIEAARAGESGSGEFAVVADEVRTIAQRTQOSTMQISEIIIEQLQEGAKNVS 527
Qy 392 DQDEVNQRIGEGVERVEEAMETLQETITDAVEDAASGMQEVSTATDEQAVSTEVEAEVMD 451
Db 528 RSMDESKLETDIVVEKTNQVNEKISLVQQAHRISDMNLQIASAAEQSLVAEEINNTV 587
Qy 452 GVDDRAGEIAAALDDIADATDQQVTVVEEVRETVGK 487
Db 588 NIKDLSIKLSEASNAGTEMNAQVSKVKBQNELINE 623

Search completed: August 10, 2004, 15:26:13
Job time : 86.6574 secs

2012 K

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:23:41 ; Search time 24.8644 Seconds
(without alignments)
1015.311 Million cell updates/sec

Title: US-09-455-978B-2
Perfect score: 2394
Sequence: 1 MSNDNDTLVADVRNGIDGH.....ATDQQVTVBVRVTVGKLS 489

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pap.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pap.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pap.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pap.*
5: /cgn2_6/ptodata/2/iaa/6C COMB.pap.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	412.5	17.2	640	4	US-09-252-991A-23252
2	401.5	16.8	545	4	US-09-252-991A-31086
3	398	16.6	663	4	US-09-252-991A-23255
4	397.5	16.6	906	4	US-09-252-991A-32715
5	389.5	16.3	696	4	US-09-252-991A-16965
6	385	16.1	710	4	US-09-252-991A-32789
7	380.5	15.9	734	4	US-09-252-991A-30703
8	380.5	15.9	857	4	US-09-252-991A-23956
9	379	15.8	701	4	US-09-252-991A-23288
10	373.5	15.6	709	4	US-09-328-352-5172
11	373	15.6	614	4	US-09-252-991A-31412
12	368.5	15.4	573	4	US-09-252-991A-18744
13	366	15.3	760	4	US-09-252-991A-31724
14	361	15.1	613	4	US-09-252-991A-25899
15	359	15.0	579	4	US-09-543-681A-6665
16	358.5	15.0	611	4	US-09-252-991A-20097
17	352.5	14.7	563	4	US-09-252-991A-31048
18	348.5	14.6	684	4	US-09-252-991A-28604
19	346.5	14.5	504	4	US-09-252-991A-26180
20	343.5	14.3	537	4	US-09-252-991A-20929
21	334.5	14.0	653	4	US-09-252-991A-18264
22	334	14.0	572	4	US-09-543-681A-8138
23	333	13.9	645	4	US-09-252-991A-16799
24	329.5	13.8	548	4	US-09-252-991A-23147
25	321.5	13.4	535	4	US-09-543-681A-4593
26	320.5	13.4	670	4	US-09-252-991A-26867
27	315	13.2	595	4	US-09-543-681A-6908

28	313	13.1	552	4	US-09-543-681A-8191	Sequence 8191, Ap
29	313	13.1	680	4	US-09-252-991A-26639	Sequence 26639, A
30	292.5	12.2	531	4	US-08-976-063E-34	Sequence 34, Appl
31	288	12.0	519	4	US-09-543-681A-6505	Sequence 6505, Ap
32	287.5	12.0	515	4	US-09-543-681A-6739	Sequence 6739, Ap
33	270.5	11.3	485	4	US-09-252-991A-24246	Sequence 24246, A
34	263	11.0	452	4	US-09-252-991A-31873	Sequence 31873, A
35	261	10.9	472	4	US-09-543-681A-6560	Sequence 6560, Ap
36	239.5	10.0	2310	4	US-09-874-923-120	Sequence 120, App
37	231.5	9.7	454	4	US-09-252-991A-28633	Sequence 28633, A
38	221.5	9.3	107	4	US-09-252-991A-21242	Sequence 21242, A
39	205	8.6	955	2	US-08-428-414A-3	Sequence 3, Appli
40	202.5	8.5	955	1	US-08-006-676B-1	Sequence 1, Appli
41	202.5	8.5	955	1	US-08-282-845-2	Sequence 2, Appli
42	202.5	8.5	955	5	PCT-US94-00324-1	Sequence 1, Appli
43	195.5	8.2	1972	4	US-08-875-435B-4	Sequence 4, Appli
44	194	8.1	1786	3	US-08-973-462-8	Sequence 8, Appli
45	191.5	8.0	1972	4	US-08-875-435B-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-252-991A-23252
; Sequence 23252, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23252
; LENGTH: 640
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23252

Query Match	17.2%;	Score	412.5;	DB	4;	Length	640;
Best Local Similarity	24.4%;	Pred. No.	3.3e-24;				
Matches	149;	Conservative	97;	Mismatches	165;	Gaps	19;
QY	7	TLVTADVRNGIDG-----HALADRIGL-----DEAEIAWRISFTGIDDDTWAALAAEQPLF	57				
Db	59	TLTASNIQSWLEGRMHVLEGLASQLALLDQPDENAIARQL-----EQPVF	103				
QY	58	-----EATADALVTDFYDHL-ESYE-RTQDLFANS-----TKTVEQ	91				
Db	104	SRNPASVYLGEASGFTMPDYDAMPEGYDPRTRAWYKDALAADRLIVTEFFVDAGTGEQ	163				
QY	92	-----LKEAQEYLLGLGRGEYDTE-----YAAQARIGKI--HDVLGL-	128				
Db	164	ILAMSLPVRHAGQLLVAAAGDKMLETTAIIINSILKFDGAGYAFVLSAGKILLHPDPSGLV	223				
QY	129	-----GPDVYLGA-----TRYTGLLD-----	146				
Db	224	LKTLAEAYPKGAPNIVPGVHEVELDGRSQVSTPVPKGLPGVTWYVALVLDRTATYSMLS	283				
QY	147	-----ALADDVVA-----DRG-----EAAAAADELVARELPMLKLL	178				
Db	284	EFTSAIVATLIAVVGIMLLGLMLIRVLMPQLTDMGRAMQDIAQEGEDLTKR-----LKVT	339				
QY	179	TFDQQLAMDTYIDSYAQRHLHDEISRQELANAVATHVAPLSLEATSQDVAERTDTMRA	238				
Db	340	SNDEFGTTLANAFNRFRVERIHESI---REVAGTA-----RQJHDVAQLVVNASNSMA	388				


```

: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252.991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 32715
: LENGTH: 906
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32715

```

[illegible]

```

RESULT 5
US-09-252-991A-16965
; Sequence 16965, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16965

```

```

; LENGTH: 696
; TYPE: prt
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16965

```

Query Match	16.3%;	Score 389.5;	DB 4;	Length 696;
Best Local Similarity	25.3%;	Pred. No. 2.3e-22;		
Matches 131; Conservative	98;	Mismatches 210;	Indels 79;	Gaps 14;

QY	1	MSNDNDTLTVADVRNGIDGHALADRI--GLDEABEIANWRLSFTGIDDDTMAALAEQPLFE	58
Db	213	LAGDENSVOAAD-SFGRDA-SLFGVLKMGMEGNAAMSISKV-TNAEAVDRLNEIAELFE	269
QY	59	ATADALVTDFYDLHLSYERTODLF-----ANSTYTVQLKETQAEYLL-----GLGGEYD	109
Db	270	-----FVSGSVD--EILETSPDLQVREAAANNIEFSVQTLDDKASQLADGFENLAGRSI	322
QY	110	TEYAAQARARIGIKHIDVLGLGPDVVLGAVTRYTG-----LLDALAD	150
Db	323	NUFA-----GVVLGALALASILLGLVWVRETNRRLAETAENKDRNQAAILLRLEIDAD	376
QY	151	DVVDARGEEAAAANDVELVARELPMLKLLTFDQIQIAMDYIIDSYAQRHLHDEIDSRQELA--	208
Db	377	--LADGDLTVAAATVTE-----DFTCALADTSINYSIDQLRELVEIT	413
QY	209	-NAVATHVEAPLUSSLEATSQVABERTDTMRARTDDQVDRMADVSREISSVSASVEEVAST	267
Db	414	INQTVAVOAAAAQETQSTAMHLAE-----ASEHQAEIAGASAAINEMAVSIDQVSAN	466
QY	268	ADDVRTSEDAEALAAQOEAAADALATWTDIDBATDGTGTAGVEQLGERAADVESVTGVI	327
Db	467	ASESSAVERSAVIAKNGNEVHHNTITGMONIREQIQDTSKRIKLGSQEQEIGDVTSLI	526
QY	328	DDIAEQTNMLNALSIEAARAGEGEGFVAVDEVKALAEESREQSTRVEELVEQMOAET	387
Db	527	NDAIDQTNILAINRAIQASMGADGARGFVAVDEVQRLAERSAAATQIEALUKTJOTDT	586
QY	388	EETVDQLDEVNQRIGEGVERVEEAMETLOEITDAVEDAASCMQEVSTATDTEQAVSTEEVA	447
Db	587	NEAVISMEOQTTSVVRGARLAQDAGVALEIEKVKSKTLAALIQNISNAAPQAQASSAGHLS	646
QY	448	EMVGDVDRAGEIAAALDDIADATDQQVTVVEEVRETV	485
Db	647	NTMNYIQBITQTSAGTTTATARSICNLAKMASERNV	684

```

RESULT 6
US-09-252-991A-32789
; Sequence 32789, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32789
; LENGTH: 710
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32789

```

Query Match 16.1%; Score 385; DB 4; Length 710;
Best Local Similarity 27.8%; Pred. No. 5.4e-22;
Matches 131; Conservative 83; Mismatches 215; Indels 42; Gaps 10;
QY 39 FTGIDDDTMAALAEOPLEAT-----ADALVTFDYDHLSEYRTODILFANS--- 85

```
Db 257 PERSDKAQAFAAFADALRQATTLRGQLPGEADAALAEQAMSGLQYRGGEQFAGVIR 316
Qy 86 TKTVQLKETQAEYLLGLGR-----GEYDTEYAAQARIGK-IHVDVLGLGPDVYLG-AYT 138
Db 317 TRQAQAMQSTQDMARAGRTLTEAGRQLRESTASDRASLWLIALALAFGCVAGWAIN 376
Qy 139 RYVTGLLD-ALADDVADRGEAAAADELVARFLPMLKLTFFOOIAMDTYIISYAQRL 197
Db 377 RQIVRPLDEALA-----QAEATAAGDLGKRPONPLTLORRDEL-----GQLQRV 420
Qy 198 HDET-DSRQELANAVATHVEAPLSLEATSDQVAERTDTMRARTDQVDRMADVSREISS 256
Db 421 MORMGDSRLRELVRIGDGV-----SOLASSAELSAVTEQTRAGVNSQKVETDQVATAMHE 476
Qy 257 VSASVEEVASTADDVRRTSDEBALAQOGEAAADALATMTDIDBATGVTAGVEQLGER 316
Db 477 MAATVQDVARNAEALSAQAARQADEARQGDVQAVTRIERLASEMDVVSSEAMARKNE 536
Qy 317 AADVSTVGVDDTARQTNMLALNASIEAARAGEGFAVVADRVKALAEESRQSTRV 376
Db 537 SEQIGSVLDVTKSVAEQTNLALNAIEAARAGDAGRGFAVVADRVGLAORTQOSTAEI 596
Qy 377 BELVQMOAETEETVDQLDEVNORIGEGVERVEEAMETLQETDADVEDAASGMQEVSTAT 436
Db 597 EGLIQRLOQAGEAAERLENSRSLTASTVELARRAGALDSITRTVSDIQNMNLQIATAA 656
Qy 437 DEQAVSTVEVAMVDGVDDRAGEIAAALDDIADATDQOVRTVVEEVRETVGK 487
Db 657 EQQSTVABEINRSLVSRVDVABQSAASSEQTAASSGELARLGTQLQAQVGR 707

RESULT 7
US-09-252-991A-30703
; Sequence 30703, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30703
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30703

Query Match 15.9%; Score 380.5; DB 4; Length 734;
Best Local Similarity 24.7%; Pred. No. 1.3e-21;
Matches 119; Conservative 91; Mismatches 178; Indels 93; Gaps 13;

Qy 41 GIDDDTMAALAAEQPLFEATADALVTFDYDHLESY-----ERTQDLFANSTKTV 90
Db 310 GNEVDNLKMLTVDPQLYDIDAE-----HGHIELFUPFTIADSGVRWTLMLQIPQAAVRG 363
Qy 91 QLKETQAEY-----LLGLGRGEYDTEYAAQARIGKTHDVLGLGPDVYLGAY-----T 138
Db 364 ELQQLQGLSDORQQDILGM-----SLAGLVAAALGL-LVVNLVGYGIARPL 409
Qy 139 RYVTGLLDALAD---DWDADRGEAAAAVDEL--VARFLPMLKLTFFOOIAMDTYIDSY 193
Db 410 RQIVGLMDDIAQEGDGLTRLSSERA---DELGSIAK-----GNFTFLGKL 452
Qy 194 AQRHDEIDSRQELANAVATHVEAPLSLEATSDQVAERTDTMRARTDQVDRMADVSRE 253
Db 453 QNMIGVQVQSVQKVS-----DSSEHTADIAIRTNQGVQOQL---AE 490
```

```
Qy 254 ISSYSASVEEVASTADDVRR-----TSEDAEALAQOGEAAADDALATMTDIDEATDGV 306
Db 491 IELVATVHMTATQDVARNATHAAEAANHADQAAGKQIVSSSSRAIQALASEIGRA 550
Qy 307 TAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEGAFVAVDEVKALA 366
Db 551 VGVVONLAKSENINAILVAIRGIAEQTNLALNAATEAARAGEQGRGFVAVDEVNLA 610
Qy 367 BESRQSTRVBELVEBQMAETEETVDQLDEVNORIGEGVERVEEAMETLQETDADVEDAA 426
Db 611 QKTOQATEEIQSMTCQLQGGTRDVVKVMQDSQERTDSDVRHARQAAEALESITQAVSVIN 670
Qy 427 SGMEVSTATDQAVSTVEEVAEMVDGVDDRAGEIAAALDDIADATDQOVRTVVEEVRETVG 486
Db 671 DMNFOIASAAEQSAVEDINRNVANICQVANOVAGGADEASQASAEITRLAEQORRLVN 730
Qy 487 K 487
Db 731 Q 731

RESULT 8
US-09-252-991A-23956
; Sequence 23956, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23956
; LENGTH: 857
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23956

Query Match 15.9%; Score 380.5; DB 4; Length 857;
Best Local Similarity 26.8%; Pred. No. 1.6e-21;
Matches 141; Conservative 92; Mismatches 226; Indels 67; Gaps 16;

Qy 2 SNNDTLVTADVNRNGIDGHALADRIGLD-----EAEIAW--RLSFTGIDD-----DTM 47
Db 352 ANSLDILQSSLAEDG--AVALKESQGHGDEPLLLQQAQVAQVNRVLVQLGLDEARSRL 409
Qy 48 AALAAEQP---LFEA-----TADALVTFDYDHLESYERTQDLFANSTKTVBOLK 93
Db 410 ADAQEGPFSRLREALELAARLEQAITDDAVYVVKVDVLTNIRGFADKLAE--YRASQJQ 467
Qy 94 ETQAEYLLGLGRGE---YDTEYAAQARIGKIHVDVLGLGPDVYLGAYTRYTGLLDALA 149
Db 468 EQMTAAMGERAGQVAMRVDRSWEAQOQAM--LHS-LRTNSLLIIVGA-----AVLALLV 518
Qy 150 DDVADRGEEAAAAVDELVARFLPMLKLTFFOOIA---MDTYIDSYAQRHLHDEIDSRQE 206
Db 519 -----GLGAFAFGISLLIVR--PLRQAMGVAHRIAEGLDAVRVDS---ERRDEVGQLMA 566
Qy 207 LANAVATHVEAPLSLE-----ATSDQVAERTDTMRARTDQVDRMADVSREISSYSA 259
Db 567 AMRAMTGLSRGTVSQDQGVGRIAGASEALSQVTRTRTRIGIDSGRAETEQTAVAMNQMAA 626
Qy 260 SVEEVASTADDVRRTSDEBALAQOGEAAADALATMTDIDEATDGVTAGVEQLGERAAD 319
Db 627 TVHEVAHNABEAAAGAESADGKVSQGVQEVVTRTLERIERLAEAVRAATASVEALSADSOR 686
Qy 320 VESVTGVIDDIAEQTNMLALNASIEAARAGEGFAVVADRVKALAEESRQSTRV 379
```


Db 207 RQRF-----LEAGADPIAGQAVGTIDRATTAAQMALRDELHQASDLHSSISAEARRT 260
QY 157 -----GEEAAAADVE--LVARFL--PMLKLTFFQQTAMDTYIDSYQRLHDEIDS 203
Db 261 MLLGSLVLIGASLAVALLSLMLVNRNLVRPQVRLIEHIAQLSHGDFGERIEIRRKDELCK 320
QY 204 RQELANAVATHEAPLSLEATSQDVAERTDT-----MRARTDQVDRMADVSRIS 256
Db 321 LALAANTURDLVDFLRRLSTRDLSDASGLNIAISLMAAGTREQRSRDTQVATAMQE 380
QY 257 VSASVEEVASTADVRRRTSEDAEALAAQGEAAADALATMTDIDEATDGTAGVQLGER 316
Db 381 MSATAQEVARYAGDAARAADDEADDSQORGEDVMEETIRSIGEMRKEIDHTVEVIRQLSD 440
QY 317 AADVESVTGVDIDIAEQTNMLNALSIEAARAGEGFAVVADVKALAESRQSTRV 376
Db 441 SGRIGKVLVDVIRGTAEQTNLLALNAAIEAARAGDAGRGFAVVADVVRTLAQRTAESIAE 500
QY 377 EELVBMQAETEETVDQDEVNQRIGEGVERVEEAMETLOETITDAVEDAASGMQEVSTAT 436
Db 501 HQIDITVONGAVNAARATIESGGRSEAGAEQVANAGAMLQRTASVESIRDMNRQIATAA 560
QY 437 DEQAVSTEEVAMVDGVDRAGEIAAALDDIADATDQQVTVVEEVRETVGKLS 489
Db 561 BEQTAVAEIDISRNLT-----EIAS-----IASSNOEQVEQTEAASRDHLGLS 602

RESULT 12
US-09-252-991A-18744
; Sequence 18744, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18744
; LENGTH: 573
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18744

Query Match 15.4%; Score 368.5; DB 4; Length 573;
Best Local Similarity 25.0%; Pred. No. 7.9e-21;
Matches 129; Conservative 89; Mismatches 175; Indels 123; Gaps 15;
QY 14 RNIGDGHALADRI-GL--DEABIAWRLSFTGTDITMAALAAEQPLFEATAD-ALVTDIFY 69
Db 100 RSLLEGHLSAEQIAGLRNDHAENLQAAL-----DRVQYAAQPF--GEAEFLAKVAEP- 149
QY 70 DHLESYETQDLFANSTKTVQELKETQAEYLLGLRGEGYDTEYAAQPARIGKIHDVLGLG 129
Db 150 -----ERGYALWSATSARVLSLAASDFSAQAQLSYGSDRQFGAMREVINQL----- 196
QY 130 PDVYLGAVTRYTGLLDALADVDRGEEAAAADVELVARFL-----PMLKLTTFDQ 183
Db 197 -----DEMEEAATAADGASGALGERHWRQOVALVAFGLL 231
QY 184 IAMDTY-----IDSYAQRLHD-----EIDSRQE-----LANAVATHV 215
Db 232 VCLSLVLPFGVGLVTRPLQRLQRLIEIANGDGLRVLRLVLTSTRDEPGRIGLSAFNAFLDKL 291
QY 216 EAPLSLLEATSQDVAERTDTMPARTDDQVDRMAD-----VSREISSVSASVEEVAST 267
Db 292 QPLIREVGRVGTGEVADSAGSLAGMTAAN--DRLINSEHASVDPQVSTAATQMSSAVHEVARN 350

QY 268 ADDVRRTSEDAAEALAAQGEAAADALATMTDIDEATDGTAGVQELGERAADVESVTGVI 327
Db 351 AQSAAQVADDDARRQAREGANVVEATIEVIRQLAQEVSSSESIQQLAQETASIDAVLTVI 410
QY 328 DDIAEQTNMLNALSIEAARAGEGFAVVADVKALAESRQSTRVEELVEIQMQAET 387
Db 411 KGIAEQTNLLALNAAIEAARAGEGQGRGFAVVADVKALAESRQSTRVEELVEIQMQAET 470
QY 388 BETVDQLDEVNQRIGEGVERVEEAMETLOETITDAVEDAASGMQEVSTATDEQAVSTEEVA 447
Db 471 QNAVAMOGSLKARDSVER-----AAGVDGVLAAT-----G 502
QY 448 EMVDGVDRAGEIAAALDDIADATDQQVTVVEEVRE 483
Db 503 DAVGRINDLAAQIASACEEQSRVIDEIAARNISEVRE 538

RESULT 13
US-09-252-991A-31724
; Sequence 31724, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31724
; LENGTH: 760
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31724

Query Match 15.3%; Score 366; DB 4; Length 760;
Best Local Similarity 31.7%; Pred. No. 1.8e-20;
Matches 97; Conservative 61; Mismatches 126; Indels 22; Gaps 4;
QY 195 QRL--HDEIDSRQELA---NAVATHVEAPLSLEATSQDVAERTDTMPARTDDQVDRMADV 250
Db 454 QRLPHTGDELGELAGWENRFLDKLQPIIRDVKSVDRDARSTADQSAISSQTSAGMQQQ 513
QY 251 SREISSVSASVEEVASTADVRRRTSEDAEALAAQGEAAADALATMTDIDEATDGTAG- 309
Db 514 FREIDQVATASHMTATADQVARSAAQAADAARGADQATRDGLALIDRTTQSIDSLSAANL 573
QY 310 -----VEQLGERAADVESVTGVDIDIAEQTNMLNALSIEAARAGEGFAVVADVK 363
Db 574 TSAMQGVBEQLASSSEESIGSVLEIRATAEQTNLLALNAAIEAARAGDAGRGFAVVADVK 633
QY 364 ALAESRQSTRVEELVEIQMQAETEETVDQDEVNQRIGEGVERVEEAMETLOETITDAVE 423
Db 634 NLARTQSVQIRGVIEGLQGTTRDVVDAMHGHRQAGSVEQVDEAVVALQRICEAVT 693
QY 424 DAASGMQEVSTATDEQAVSTEEVAMVDGVDRAGEIAAALDDIADATDQQVTVVEEVRE 483
Db 694 VINDMNLQIASAAEQSSVAEEINRNV-----AAIRDVTESLSQAEESAQVSQ 742
QY 484 TVGKLS 489
Db 743 SLNRLA 748

RESULT 14
US-09-252-991A-25899
; Sequence 25899, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

```

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25899
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-25899

```

Query Match	15.0%;	Score 359;	DB 4;	Length 579;
Best Local Similarity	25.9%;	Pred. No. 4.4e-20;		
Matches 133;	Conservative 85;	Mismatches 181;	Indels 114;	Gaps 15;
QY	44	DDTMAALAAEQPLFEATADALVTDVFDHLESY----	ERTQDLFANSTKTVSQLKETOAYEL	100
DB	70	DNTYAAU-----NLRITDIIAQAFSYIADPOVED--	ANITTFIQTMDARKA----	113
QY	101	LGLGRGEYDT---EYAAQARARIKIHVDVLGLGPDVILGAYTRYTYGLL-----		145
DB	114	-----EVDTLMNEYVELSAQTG--PDQORMAEIKQLYQKSRVDLDELVOYLNERNTNAI		165
QY	146	-----DALADDVVDAGBEGAAAAVDEL-----	VARFLPMKLKLL	178
DB	166	RLIKSPANDFNFTNSLYEATDFTDDEVNPSVTEAEDNYHEMLWIASLFGMIFLITFTLV		225
QY	179	TF-----DQQLAMDYTIOS--YAOQLRDEIDSRQEL-----	ANAVATHVPEAPLS	220
DB	226	LFWIRKHIIIVRNQMIYQBAISGDGLISRIDHDTGRIEIDQLMLGLOQMRARKKEWVS		285
QY	221	SLEATS-----QDVAERTDTTMRATDDQVDRMADVRSREISSVSASVEEVASTADDDVRR		273
DB	286	AIRNSSTIISYGVQIEAAGNDLSSRTEEQASALEETASSMEQLTAIVRNNTESAREVTQ		345
QY	274	TSDAEALAAQCGEAAADDALATMTDIDEATGDGTGVAQVEQLGERAADVESVTGUIDDIAEQ		333
DB	346	LVMSTADIAVQGGEGHSKNWMTMTDI-----	ADRSQKIGEITAVTIDDIAFQ	391
QY	334	TNMLALNASIEAARAGEAGFVAVDEVKALAAEESREQSTRVEELVEQMQAEETEETVDQ		393
DB	392	TNIALNAVAEERAGEQGGEGFSVVAEVENLRALQRSAAEAKIKELIESTILRVKQNGDL		451
QY	394	LDEVNQRIGEGVERVEEAMETLBOITIDAVBDAAASGMQEVSTATDEQAVSTEVEAEMDGV		453
DB	452	VEQVLSMGEIVTISVNVHVSMSMKELLAASEEQRTGIAQISLAVNEMDKATQONAAW----		508
QY	454	DDRAGETAAALDDIADATDQOVRT--VEEVRET		484
DB	509	EQSTAASATLSSEQASMLDMIVNTFKVEQAVT		540

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:28:46 ; Search time 71.0412 Seconds
(without alignments)
2159.184 Million cell updates/sec

Title: US-09-455-978B-2

Perfect score: 2394

Sequence: 1 MSNDNTLTADVRNGIDGH.....ATDQQRVTEVEVETVKLS 489

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	399.5	16.7	535	12	US-10-282-122A-66393
2	399.5	16.7	535	12	Sequence 66393, A
3	386.5	16.1	682	12	Sequence 531, App
4	374.5	15.6	680	12	Sequence 66174, A
5	371.5	15.5	686	12	Sequence 69670, A
6	361.5	15.1	644	12	Sequence 68162, A
7	359.5	15.0	891	9	Sequence 77591, A
8	356.5	14.9	539	12	Sequence 5, Appli
9	348.5	14.6	679	12	Sequence 51025, A
10	348	14.5	626	12	Sequence 372, App
11	338	14.1	686	12	Sequence 77066, A
12	332	13.9	547	12	Sequence 52305, A
13	326	13.6	641	12	Sequence 28, Appl
14	315	13.2	643	12	Sequence 77278, A
15	314.5	13.1	501	12	Sequence 77602, A
					Sequence 69564, A

16 313.5 13.1 541 12 US-10-282-122A-77047
17 308 12.9 845 12 US-10-282-122A-77047
18 292.5 12.2 531 8 US-10-282-122A-76517
19 292.5 12.2 531 11 US-09-750-986D-34
20 292 12.2 564 12 US-10-335-977-6156
21 289 12.1 564 12 US-10-335-977-6157
22 289 12.1 565 10 US-09-882-227-462
23 284 11.9 630 10 US-09-882-227-460
24 280 11.7 675 12 US-10-335-977-4944
25 276 11.5 293 12 US-10-335-977-4942
26 273.5 11.4 664 12 US-10-335-977-4943
27 269 11.2 753 12 US-10-282-122A-47230
28 263 11.0 431 12 US-10-389-647-451
29 257 10.7 431 12 US-10-012-819-228
30 254 10.6 433 8 US-08-945-038-6
31 252.5 10.5 433 12 US-10-335-977-8500
32 252.5 10.5 438 12 US-10-335-977-8501
33 245 10.2 883 15 US-10-369-493-18563
34 241 10.1 654 12 US-10-282-122A-76558
35 241 10.1 673 12 US-10-335-977-6249
36 239.5 10.0 2310 9 US-09-874-923-120
37 239.5 10.0 2310 9 US-09-991-496-120
38 239.5 10.0 2310 10 US-09-820-843A-114
39 231.5 9.7 2354 10 US-09-820-843A-113
40 225 9.4 1190 15 US-10-369-493-18546
41 221 9.2 566 12 US-10-243-552-921
42 217.5 9.1 202 12 US-10-335-977-6246
43 212 8.9 1258 14 US-10-156-761-10395
44 212 8.9 2124 15 US-10-369-493-11841
45 209 8.7 533 12 US-10-282-122A-49144

ALIGNMENTS

RESULT 1

US-10-282-122A-66393
; Sequence 66393, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636

```
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66393
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-66393

Query Match      16.7%; Score 399.5; DB 12; Length 535;
Best Local Similarity 25.2%; Pred. No. 1.8e-18;
Matches 132; Conservative 111; Mismatches 206; Indels 75; Gaps 16;

QY 3 NNDTTLVTADVNRNGIDGHALADRLGDEAEIAWRLSFTGIDDDTMAALAAQPLFEATAD 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 37 SENELSVNA-LRNHMEGDMHD-----ALRADVLAAFWV-QPGDGAAG 78

QY 63 ALVTDFVHLESYERTQDLFANSTKTVEQ-----LKETQAEYLLGLGRGYDTEYAAQRA 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 79 QVRDQLQHSQWFR-----KVVEQNOGLPLNDIAHQALVEL-RPDLEAVIGAAES 127

QY 118 RIGK-IHDVLGLGPDV--YLGAYTRYTTGLLDALADDV-----VADRGEE----- 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 128 IVGKALLDPVAARAELPQFVQAF-KELEGRNEALSSLIEKHVEQTNRAEDSMRYSAWML 186

QY 160 -----AAAADVELVARPL-----PMLKLLTFDQOIAMDTY-----IDS--YAQRLHDEI 201
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 187 AGGILVACLVLGQLCQLLRAVLQPLRLKLVSSARVIAQGNLQEPITGVDSNDEAGLQAL 246

QY 202 DSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQVDRMADVSEISSVSASV 261
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 247 GEMENLRQMTITIRQSEELHDTSSQSGTQSVHVGASQQADSAATSMASMEEMITNI 306

QY 262 EVASTADDVVRTSEDAEALAQOGEAAADALATMTDIDEATDGTAGVEQLGERAADVE 321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 307 SQISDHADNARVISAKSEELASSGGQVILNVVEGMSRIADVNVQSSTISITAGQSSDEIH 366

QY 322 SVTGVDDIADQTNMALNASEIARAAGEGFAVVADEKALAEBSREOSTRVEELVE 381
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 367 SIIQVIRGIAEQTNLLALNAIEAARAAGEAGRFAVVADEVRLGAARTTQSTQBITAMIE 426

QY 382 QMQAETETVDQDEVNQRIGEGVERVEEAMETLQETITDAVEDAASGMQEVSTATDEQAV 441
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 427 RIRASTQAINSMAGVSRVNEGVSFAQAGVSNIEILDGTRHAASVVDEISQITREQSR 486

QY 442 STEEVAEMVGDVDRAGEIAAALDDIADATDQQVRTVEEVRETV 485
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 487 ASDEIAQRVELIAQRSQNTQAMHEMA-AT---ARRLNEVAATM 526

RESULT 2
US-10-389-647-531
; Sequence 531, Application US/10389647
; Publication No. US20040033549A1
; GENERAL INFORMATION:
; APPLICANT: GREENBERG, E. Peter
; APPLICANT: SCHUSTER, Martin
; APPLICANT: LOSTROH, Candi
; TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
; FILE REFERENCE: UIZ-038CP
; CURRENT APPLICATION NUMBER: US/10/389,647
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: 09/653730
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/153022
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 710
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 531
; LENGTH: 535
```

```
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-389-647-531

Query Match      16.7%; Score 399.5; DB 12; Length 535;
Best Local Similarity 25.2%; Pred. No. 1.8e-18;
Matches 132; Conservative 111; Mismatches 206; Indels 75; Gaps 16;

QY 3 NNDTTLVTADVNRNGIDGHALADRLGDEAEIAWRLSFTGIDDDTMAALAAQPLFEATAD 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 37 SENELSVNA-LRNHMEGDMHD-----ALRADVLAAFWV-QPGDGAAG 78

QY 63 ALVTDFVHLESYERTQDLFANSTKTVEQ-----LKETQAEYLLGLGRGYDTEYAAQRA 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 79 QVRDQLQHSQWFR-----KVVEQNOGLPLNDIAHQALVEL-RPDLEAVIGAAES 127

QY 118 RIGK-IHDVLGLGPDV--YLGAYTRYTTGLLDALADDV-----VADRGEE----- 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 128 IVGKALLDPVAARAELPQFVQAF-KELEGRNEALSSLIEKHVEQTNRAEDSMRYSAWML 186

QY 160 -----AAAADVELVARPL-----PMLKLLTFDQOIAMDTY-----IDS--YAQRLHDEI 201
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 187 AGGILVACLVLGQLCQLLRAVLQPLRLKLVSSARVIAQGNLQEPITGVDSNDEAGLQAL 246

QY 202 DSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQVDRMADVSEISSVSASV 261
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 247 GEMENLRQMTITIRQSEELHDTSSQSGTQSVHVGASQQADSAATSMASMEEMITNI 306

QY 262 EVASTADDVVRTSEDAEALAQOGEAAADALATMTDIDEATDGTAGVEQLGERAADVE 321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 307 SQISDHADNARVISAKSEELASSGGQVILNVVEGMSRIADVNVQSSTISITAGQSSDEIH 366

QY 322 SVTGVDDIADQTNMALNASEIARAAGEGFAVVADEKALAEBSREOSTRVEELVE 381
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 367 SIIQVIRGIAEQTNLLALNAIEAARAAGEAGRFAVVADEVRLGAARTTQSTQBITAMIE 426

QY 382 QMQAETETVDQDEVNQRIGEGVERVEEAMETLQETITDAVEDAASGMQEVSTATDEQAV 441
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 427 RIRASTQAINSMAGVSRVNEGVSFAQAGVSNIEILDGTRHAASVVDEISQITREQSR 486

QY 442 STEEVAEMVGDVDRAGEIAAALDDIADATDQQVRTVEEVRETV 485
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 487 ASDEIAQRVELIAQRSQNTQAMHEMA-AT---ARRLNEVAATM 526

RESULT 3
US-10-282-122A-66174
; Sequence 66174, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zvekand, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
```



```
Db 199 SFKIEDTGFVFLTNAOGEVQIHRQKEQV-----KSLQQIYSGSALLNKGSGFNLLSTDY 254
QY 113 AAQRARIGKTHDVLGLGPDVYLGAYTRYTGLL---DALAD-DVVADR---GEEAAAAYD 165
Db 255 QGEVWVASI-----YISMDFVLYGTPVHEVFAELDAVAQRMMLTLVAIAIF 304
QY 166 ELVARFLPMLKLLTFDQOIAMDYIDSYAQRLLHD-----BIDSRQELIA----- 208
Db 305 IFMIGIFLA-----NSIAMP--INQIAKRFTDLGRGDGDLQRIEVKGNDIQAOLSKG 354
QY 209 -NAVATHEAPLSLEATSQD-----VAERTDTMTARTDDQVDRMADYSRSTSSVSAS 260
Db 355 FNGFIEIKHOSIKDVAQTSRELQVAAGVSKALVTHDNSQQQQRDTIQVVTAINQMGAT 414
QY 261 VEEVASTADDVRRRTSEDAEALAOQGEAAADALATMTDIDBATDGTAGVEQLGERAADV 320
Db 415 ISEIASNATAETANQASGNADQGRNVVVKAKEAISRLAHDIENTGKVVEQLASTTQEI 474
QY 321 ESVTGVDDIDIAEQTNMLNALSIEAARAGEGFAVVADEVKALAEBSREQSTRVEELV 380
Db 475 GSILDAIRGISEQTNLLNALNAIEAARAGDQGRGFAVVADEVRLASFTASSTEEIQKMI 534
QY 381 EQMAEETVTDQDEVNQRIGEGVERVEEAMETLOEITDAVEDAASGMQBVSTATDEQA 440
Db 535 NQLQNDAKNAVSAAMDAGKTVTHQGVAADEAVQVLMISDRHLDISDRNTQVATATERQS 594
QY 441 VSTEVAEMVGVDDRAGEIAAALDDIADATDQQVTVVEEVRTVGKL 488
Db 595 TVVHTINQIEINAINEVTTSTAEELADAS-----KSLRELSGRL 635

RESULT 7
US-09-272-809-5
; Sequence 5, Application US/09272809
; Patent No. US20020022239A1
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C.
; TITLE OF INVENTION: Phytofluors as fluorescent labels
; FILE REFERENCE: 2500.118U50
; CURRENT APPLICATION NUMBER: US/09/272,809
; CURRENT FILING DATE: 1998-03-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 891
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: cph5 locus SLL0041
; OTHER INFORMATION: (locus 1001300) an 891 aa protein,
; OTHER INFORMATION: methyl-accepting chemotaxis protein I. Homology
; OTHER INFORMATION: to tsr in last 250 aa.
US-09-272-809-5
Query Match 15.0%; Score 359.5; DB 9; Length 891;
Best Local Similarity 27.1%; Pred. No. 1.5e-15;
Matches 136; Conservative 86; Mismatches 197; Indels 83; Gaps 19;
QY 23 ADRIGLDEAIEAWRLSFTGIDDDTMAALAAEQP--LFEATADALVTDYDHLSEY----- 75
Db 418 ADRV-----IVRFQATWAGTVVSVSABGYKPAKGATADPCFADSY--VEKYRSGR 469
QY 76 ERTQDLFANSTKT---VEQLK--ETQAEYLLGLGRGEYDTEYAAQARIGKIHVLGL-- 128
Db 470 QATRDYI-NAGLTPCHIGQLKPEVKANLV-----APINKGNLLGLLI 512
QY 129 -----GP-----DVYLGATRYTYTGLL-----DALADDVVADGGERAAAVDELVARFLP 173
Db 513 AHQCSGRPDWHQNEIDLFGQLTVQVGLALERSDLLAQOKIAE--VEQQRMRKMQKRALE 570
QY 174 MLKLLTFDQOIAMDTYIDSVAQRLLHDEIDSRQELANAVATHVAPLSSLEATSQDVAERT 233
```

```
Db 571 L--LMEVDPVSRGDLTIRAHV--TEDEIGTIADSYNATIESLRRIVTQVTAASQFTETT 626
QY 234 DT-----MRARTDDQVDRMADVSRFISVSASVEEVASTADDVRRTSDEALAOQGE 286
Db 627 DTNEVAVRQAOANRQALDVAELERLQAMNKSIOQVAENAAQAESAVERATQTVDOGE 686
QY 287 AAADDAALATMTDIDBATDGTAGVEQLGERAADVESVTGVDDIDIAEQTNMLNALSIEAA 346
Db 687 DAMRTVDGIVATRETVAATAKQVKRLGESSOKISKVNVNIGSFADQTNLLNALNAIEAA 746
QY 347 RAGEAGEGFAVVADEVKALAEBSREQSTRVEELVEQMAETEETVTDQDEVNQRIGEGVE 406
Db 747 HAGEGEGFAVVADEVSRARQSAEATAEIAQLVATTIQAETNEVYNVNAMEAGTEQVVVGTG 806
QY 407 RVBEAMETLOEITDAVEDAASGMQE--VSTATDEQAVSTEE-----VAEMVD----- 451
Db 807 LVEETRESLQNQIT-AVSAQISGLVEALTSAAIEQSQISESVTQTMALVAQIADKNSSEAS 865
QY 452 GVDDRAGEIAAALDDIADATDQ 473
Db 866 GVSATFKELLAVASQLOEAVKQ 887

RESULT 8
US-10-282-122A-51025
; Sequence 51025, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51025
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Bordetella pertussis
US-10-282-122A-51025
```

Query Match	14.9%;	Score	356.5;	DB	12;	Length	539;
Best Local Similarity	26.6%;	Pred. No.	1.3e-15;				
Matches	126;	Conservative	81;	Mismatches	153;	Indels	113;
Gaps	15;						
Qy	38	SFTGIDDDTMAAL-----AAQPLEFATADALVTDYFDHLSYERTQDLFANSTKTV	EQ	92			
Db	126	SFAASLDENMAALERNDAAVLQKNVKGQASAAFAARLGEFSTNLKDSSETLAAHET	185				
Qy	93	KETQAEYLLGLGRGYDEYAAQARARIKIHVDVLGLGPDVYLGYTRYTGLLDALADDV	152				
Db	186	RETIMLYV-----YAA-----LLLLIVGIAGSYLWYTRAVVRPLQR--	222				
Qy	153	VADRGEEAAAADVELVARFLPMLKILLTPDQOIAMDTYDSYAORLHDEIDSRQELANAVA	212				
Db	223	-ASQOQFERMAAGD-LSARI-----EGNSRNEIGALLT	252				
Qy	213	T--HVEAPLS-----SLEATSODVAERTDTWRATDDQVDRMADVRSIESSVA	259				
Db	253	ALRHQWESLRTVAARRGVDEINVGSRISAGNTDLSRSTEEQ-----AASLEETAA	305				
Qy	260	SVEEVAST-----ADVVRTSB--DABALAQGEAAADDALATWTDIDEATDGVTAGVEQ	312				
Db	306	SMEQLASTVKNQANDANQAQLAASADVAESGGSGAVSEVVSTN-----DGISASSRK	358				
Qy	313	LGERAADVESYTGVIDDIAEQTNMLNASTAARAGEAGEGFAVADVKALAEBSREQ	372				
Db	359	ISE-----IVSVIDGIAFTNLALNAAVEAARAGEQCKGFVAVAGEVRSIAQRSAQA	411				
Qy	373	STRVEELVEQMAETETVTQDLDEVNQRIGEGVERVEEAMETLQETDAVEDAASGMQEV	432				
Db	412	AKBKIVLIE-----DSVD-----KVGTSQQVERAGATWQEIIVASVKRYTDMGEI	457				
Qy	433	STATDEQAVSTEEVAEMVDGVDDPRAGEIAALDDIADATDQVTRTVEEVRETV	485				
Db	458	SAASEGSSGIEQVNRVMSQMDVETQQAALVEFAAAAGSLGQEQQRALAEAV	510				

```

RESULT 9
US-10-389-647-372
; Sequence 372, Application US/10389647
; Publication No. US20040033549A1
; GENERAL INFORMATION:
; APPLICANT: GREENBERG, E. Peter
; APPLICANT: SCHUSTER, Martin
; APPLICANT: LOSTROH, Candi
; TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
; FILE REFERENCE: UIZ-038CP
; CURRENT APPLICATION NUMBER: US/10/389,647
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: 09/653730
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/153022
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 710
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 372
; LENGTH: 679
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-389-647-372

```

Query Match	14.6%;	Score 348.5;	DB 12;	Length 679;
Best Local Similarity	25.3%;	Pred. No. 5.9e-15;		
Matches	123;	Conservative	78;	Mismatches 179;
			Indels	95;
			Gaps	14;
QY	20	HALADRGIDDEAI	---ANRLSTGIDDDTMAALAAEQPLFEATADAL---VTDFYDHLES	74
Db	242	HLIANLTGVHKAELNLCGRRFSLDV	-----PVFDANERLGSAAVQMTDTEE	289
QY	75	YERTQDLFANSTYTVQLKETQAEYLLGLGRGYDTEYAAQARIGKIHDVLGLGPDVYL	134	
Db	290	HRAEQE-----VSQVQAAA-----GDFSKRVBEAGKEGFFSLAKD---	337	

QY	135	GAYTRYTYTGLDALADDVWVADRGEEAAAAVDLVARFLFMLKLLTFDDQOIAMDYIDSYA	194
Db	328	-----LNSLVD--TADRLRD-----VSRMLGALAQGLTLQRIEAD-YQGTFG	367
QY	195	QRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRATDDQVDRMADVREI	254
Db	368	QLKDFSNDTAQLSRMLGQIIEA-ADTINTAASTIASGNAELSAETEOQASSLEETAGSM	426
QY	255	SSVSASVEEVASTADDVVERTSEDAEALAAQCGEEAADDALATWTDDIDEATDGTAGVEQILG	314
Db	427	EELTSTVKLNAENARQANSIAANASEVATQGGTVQKVVTWSSINES-----	474
QY	315	ERADVRSVTVIIDDIABOTNMALINASIEAARAGEAGEGPAVVADEVKALAEHSREQST	374
Db	475	--ARKIADIIGVIDGIAFQTNILNAAVEAARAGEQGRGPAVVAGEVRTLQORSAATAK	532
QY	375	RVEELVEQMAETEETVDQLDEVNORIGEGVERVEEAMETLQETLDAVEDAASGMQEVST	434
Db	533	EIKTLIS-----DSVDKVE-----NGNTLVAQAGTMSDIVAIRRTVDTIMSEIAA	578
QY	435	ATDQAVSTVEVAVMGVDDRAGEIAAALDDDIADATDQQVTVVEEVRETVGKLS	489
Db	579	ASAEQSTGIEVNSAVSQMDMTQONAAALVEEAAAA-----AEAMQEQAAGLLN	626
RESULT 10			
US-10-282-122A-77066			
; Sequence 77066, Application US/10282122A			
; Publication No. US20040029129A1			
; GENERAL INFORMATION:			
; APPLICANT: Wang, Liangsu			
; APPLICANT: Zamudio, Carlos			
; APPLICANT: Malone, Cheryl			
; APPLICANT: Haselbeck, Robert			
; APPLICANT: Ohlsen, Kari			
; APPLICANT: Zyskind, Judith			
; APPLICANT: Wall, Daniel			
; APPLICANT: Trawick, John			
; APPLICANT: Carr, Grant			
; APPLICANT: Yamamoto, Robert			
; APPLICANT: Forsyth, R.			
; APPLICANT: Xu, H.			
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms			
; FILE REFERENCE: ELITRA.034A			
; CURRENT APPLICATION NUMBER: US/10/282,122A			
; CURRENT FILING DATE: 2003-02-20			
; PRIOR APPLICATION NUMBER: 60/191,078			
; PRIOR FILING DATE: 2000-03-21			
; PRIOR APPLICATION NUMBER: 60/206,848			
; PRIOR FILING DATE: 2000-05-23			
; PRIOR APPLICATION NUMBER: 60/207,727			
; PRIOR FILING DATE: 2000-05-26			
; PRIOR APPLICATION NUMBER: 60/230,335			
; PRIOR FILING DATE: 2000-09-06			
; PRIOR APPLICATION NUMBER: 60/230,347			
; PRIOR FILING DATE: 2000-09-09			
; PRIOR APPLICATION NUMBER: 60/242,578			
; PRIOR FILING DATE: 2000-10-23			
; PRIOR APPLICATION NUMBER: 60/253,625			
; PRIOR FILING DATE: 2000-11-27			
; PRIOR APPLICATION NUMBER: 60/257,931			
; PRIOR FILING DATE: 2000-12-22			
; PRIOR APPLICATION NUMBER: 60/267,636			
; PRIOR FILING DATE: 2001-02-09			
; PRIOR APPLICATION NUMBER: 60/269,308			
; PRIOR FILING DATE: 2001-02-16			
; Remaining Prior Application data removed - See File Wrapper or PALM.			
; NUMBER OF SEQ ID NOS: 78614			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 77066			
; LENGTH: 626			
; TYPE: PRT			
; ORGANISM: Vibrio cholerae			

```
US-10-282-122A-77066
Query Match      14.5%; Score 348; DB 12; Length 626;
Best Local Similarity 27.4%; Pred. No. 5.7e-15;
Matches 92; Conservative 78; Mismatches 134; Indels 32; Gaps 6;

QY 174 MLKLLTPDQI-----AMDYIISYAQRLHDEIDSRQELANAVATHVEAPLSS 221
Db 298 MSPLKTLDSAIKDIAASGGDLTKLIDNLDKFEFSELALGFNSFTMLGSIQIRQLKTIASG 357
QY 222 LEATSQDVARTDTMRATDQDVRMADVRSREISSVSASVEEVASTADDVVRTSEDAEAL 281
Db 358 VLGAETKANEAEVSRVLVEQQL-----QELQOLATANMEMANTASEVANSQAQVADA 410
QY 282 AQCGEAAADALATMTDIDEATDGT-----AGVEQLGE--RAAD-VESVTGVDDIA 331
Db 411 AKEGESA--SLEGGSSVHHETTDALQRLSIRIGSSVEDVKELVATRIETVLDVINDIA 467
QY 332 EQTNMLAINASIEARAGEGFAVVADEVKALAEBSREOSTRVEELVEQMQAETEETV 391
Db 468 DQTNLLALNAIEAARAGESGRGFAVVADEVRTTLAQTQQSTMQISETIEQLQEGAKNVS 527
QY 392 DQLEVNQRIGEGVERVEEAMETLQETITDAVEDAASGMQEVSTATDEQAVSTEEVAMVD 451
Db 528 RSMDESKLETDIVVEKTNQVNEKISLVQQAHLIRSDMMLQIASAABEQSILVAEELNNTV 587
QY 452 GVDDRAGEIAAALDDIADATDQQVRTVEEVRETGVK 487
Db 588 NIKLSIKLSEASAGNAGTEMNAQVSKVKEQNELLNE 623

RESULT 11
US-10-282-122A-52305
; Sequence 52305, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52305
; LENGTH: 686
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-282-122A-52305

Query Match      14.1%; Score 338; DB 12; Length 686;
Best Local Similarity 26.8%; Pred. No. 3e-14;
Matches 91; Conservative 80; Mismatches 130; Indels 38; Gaps 4;

QY 178 LTFDQQTAMDTYIDSYAQRHLHDEIDSRQELANAV-----ATHVEAPLSSLE---ATSQ 227
Db 350 LTKVEIVNEDEIGKLSKIFNTMIDSLREITRNINFSIQLAGSSQELSSAEQTSASVE 409
QY 228 DVAERTDTMRATDQDVRMADVRSREISSVSASVEEVASTADDVVRTSEDAEALAQGEA 287
Db 410 EISSATEEIASGAENQVKASNESSLLMNDVMGNMYTLKEEFDEIISFSNNTNTLASKQE 469
QY 288 AADALATMTDIDEATDQGTAGVQLGERAADVESVTGVDDIAEQTNMLAINASIEAAR 347
Db 470 NMSNMVQOMATIKNSVNVSSNIMYDLOKNSEEIGNIVEIINTIADQTNMLAINASIEAAR 529
QY 348 AGEAGEGFAVVADEVKALAEBSREOSTRVEELVEQMQAETEETVDQLDENVNRIGEGVER 407
Db 530 AGEAGKFAVVADEVKRLAEESINSANNIKLIMNTQDKTKTALNSIKDGASQSEKGESI 589
QY 408 VEEAMETLQEI-----TDAVEDAASGMQEVSTATDEQAVSTEEV 446
Db 590 VAEVKESLGEILNGFSNVNHFASVDSMTASNDISITAMASKLYDIETISNTASANTEEV 649
QY 447 AEMVDGVDDRAGETAAALDDIADATDQQVRTVEEVRETGV 485
Db 650 AASTE-----EOSATIEEITESIEKLVSMVENLKESV 681

RESULT 12
US-10-332-288-28
; Sequence 28, Application US/10332288
; Publication No. US20040054165A1
; GENERAL INFORMATION:
; APPLICANT: RAINNEY, Paul Barton
; APPLICANT: SPIERS, Andrew Julien
; APPLICANT: BANTINAKI, Eleni
; TITLE OF INVENTION: BACTERIAL POLYSACCHARIDE AND BIOFILM DEVELOPMENT
; FILE REFERENCE: I0317.70012US00
; CURRENT APPLICATION NUMBER: US/10/332,288
; CURRENT FILING DATE: 2003-10-06
; PRIOR APPLICATION NUMBER: PCT/GB01/03077
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: UK 0016842.7
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
US-10-332-288-28

Query Match      13.9%; Score 332; DB 12; Length 547;
Best Local Similarity 25.0%; Pred. No. 5.6e-14;
Matches 119; Conservative 84; Mismatches 189; Indels 84; Gaps 12;

QY 66 TDFYDHLESYERTQDLFANSTKVEQLKETQAEYLLGLGRGEYDTYAA-----114
Db 89 TDKNDYKSFARIEQQMANVEKTI-----HGQAD-----RMEFDNFKAHINYNKVLAQV 138
QY 115 -QPARIGKTHDVULGLGPDVILGAYTRYTYGLLDALADDV-VADRGEEAAAADVDELV--- 168
Db 139 LERVEANDLPGANQLLEEQLTPIWTEGRMKLNDIITENKNVSDR---ATAADEAVLSAK 195
```

```

, FILE REFERENCE: ELITRA.034A
, CURRENT APPLICATION NUMBER: US/10/282,122A
, CURRENT FILING DATE: 2003-02-20
, PRIOR APPLICATION NUMBER: 60/191,078
, PRIOR FILING DATE: 2000-03-21
, PRIOR APPLICATION NUMBER: 60/206,848
, PRIOR FILING DATE: 2000-05-23
, PRIOR APPLICATION NUMBER: 60/207,727
, PRIOR FILING DATE: 2000-05-26
, PRIOR APPLICATION NUMBER: 60/230,335
, PRIOR FILING DATE: 2000-09-06
, PRIOR APPLICATION NUMBER: 60/230,347
, PRIOR FILING DATE: 2000-09-09
, PRIOR APPLICATION NUMBER: 60/242,578
, PRIOR FILING DATE: 2000-10-23
, PRIOR APPLICATION NUMBER: 60/253,625
, PRIOR FILING DATE: 2000-11-27
, PRIOR APPLICATION NUMBER: 60/257,931
, PRIOR FILING DATE: 2000-12-22
, PRIOR APPLICATION NUMBER: 60/267,636
, PRIOR FILING DATE: 2001-02-09
, PRIOR APPLICATION NUMBER: 60/269,308
, PRIOR FILING DATE: 2001-02-16
, Remaining Prior Application data removed - See File Wrapper or PALM.
, NUMBER OF SEQ ID NOS: 78614
, SOFTWARE: PatentIn version 3.1
, SEQ ID NO 77278
, LENGTH: 641
, TYPE: FRT
, ORGANISM: Vibrio cholerae

```

RESULT 14

US-10-282-122A-77602

Sequence 77602, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Esse

FILE REFERENCE: ELIPRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347


```

; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77602
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-10-282-122A-77602

Query Match      13.2%; Score 315; DB 12; Length 643;
Best Local Similarity 24.9%; Pred. No. 9.2e-13;
Matches 100; Conservative 77; Mismatches 167; Indels 58; Gaps 10;

QY 133 YLG-AYTRYTGLDADLADD-----VVADGEEAAAADDELVAR----- 170
Db 249 YLGIEFGDYKRV--NLTDMMYTLMVFDKA--TALADVDVAVANAITTGIVLIGISVG 305
QY 171 -----FLPMLKLLFDQIQIAMDYIDSVAQRHDEIDSRQELAN----- 209
Db 306 IIFIINQIKPLRLKLVAVDLAQS--GDUTRRU--EVNGNDLAIQISEGFNFSGNLQ 361
QY 210 ----AVATHVEAPLSSLEATSQDVAERTDTMTARTDDQVDRMADVRSISVSASVEVA 265
Db 362 NMMLQISDATQIISSSIQQLSQ-TAKENEQMLISHSETDQ---VVTALTQMSSESAR-TVA 417
QY 266 STADDVRTSDEAALQAQGBAADDALATMTDDIDEATDGTAGVEQLGERAADVESVTG 325
Db 418 ESVTQSNQITAAKSAEQSLVIVNAVSTVTSLVNVDENMSSESISNNNRDANKISEVLS 477
QY 326 VIDDITARTNMLANASTAARAGAGGFAVVADEVKALAESEQSTRVEELVEQMOA 385
Db 478 VIGALSEQTNLIALNALEAARAGQGRGFAVVADEVKALARTQNSTEISDMLTKLE 537
QY 386 ETEETVDQLDEVNQRIGEGVERVEEAMETLQBITDAVEDAASGMQEVSTATDEQAVSTEE 445
Db 538 GTDSVVGAMERTKQCQTADTKTSEVSGSLNMSASVSDDIDLSTQIAAATQEQSTVAE 597
QY 446 VAMVVDGVDDRAGEIAALDDIADATDQOVRTVEEVRETGVK 487
Db 598 LSRNMLSIRESLVSVSGRTVVGATESLSHNSHLEQLVGK 639

RESULT 15
US-10-282-122A-69564
; Sequence 69564, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
```

```

; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69564
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Pseudomonas syringae
US-10-282-122A-69564

Query Match      13.1%; Score 314.5; DB 12; Length 501;
Best Local Similarity 27.4%; Pred. No. 7.3e-13;
Matches 86; Conservative 59; Mismatches 136; Indels 33; Gaps 5;

QY 176 KLLTFDQIQIAMDTYID-----SYAQRHDEIDSRQELANAVATHVEAPLSSLEATSQD 228
Db 184 RMLATTNQFSPDSRIDERSGKHSVLAQRFEQFLAQITGLVDGVVRDTR----GLGELGHD 239
QY 229 VAERTDTMR-----ARTDDQVDRMADVRSREISS-VSASVEEVASTADDVVRTSED 277
Db 240 LAKASGTLETGAQHQLSEIARMTGAMQRMGDMNDISGHVAVQVRAGDASDV----- 293
QY 278 AEALAAQCEAAADALATMTDIDEATDGTAGVEQLGERAADVESVTGVIDDIAEQTNML 337
Db 294 -----AHGRDSVDRAQSEITQLAARISTTDETVOALANQSEIQIGKVLVDVIGSIAEQTNLL 348
QY 338 ALNASTAARAGEGEGFAVVADEVKALAESEQSTRVEELVEQMOAETETVVDQLDEV 397
Db 349 ALNASTAARAGEQGRGFAVVADEVKALAESEQSTRVEELVEQMOAETETVVDQLDEV 408
QY 398 NQRIGEGVERVEEAMETLQBITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRA 457
Db 409 LQGVGRCVENSQRASELSRVSVEGEGICHITQLNGLIATTEQQATASREIADQLRSVQAIA 468
QY 458 GEIAAALDDIADAT 471
Db 469 EHTAANIGVIASSS 482

Search completed: August 10, 2004, 15:42:26
Job time : 73.0412 secs
```

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:22:45 ; Search time 21.3123 Seconds
(without alignments)
2207.061 Million cell updates/sec

Title: US-09-455-978B-2

Perfect score: 2394

Sequence: 1 MSNDNDTLVTADVVRNGIDGH.....ATDQOQRTVEVRETVGKLS 489

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2394	100.0	489	2 T44978	transducer protein
2	2360	98.6	489	2 B84304	Htr10 transducer [
3	643	26.9	481	2 A84294	Htr9 transducer [i
4	609.5	25.5	482	2 T44973	transducer protein
5	597	24.9	633	2 H84336	Htr3 transducer [i
6	597	24.9	805	2 T48840	transducer protein
7	580	24.2	790	2 B84238	Htr18 transducer [
8	578	24.1	778	2 F84237	Htr4 transducer [i
9	573	23.9	778	2 T48897	transducer protein
10	560.5	23.4	452	2 T44849	Htr5 transducer [i
11	555	23.2	810	2 F84327	halobacterial tran
12	554	23.1	810	2 T46810	transducer protein
13	540.5	22.6	777	2 T44597	sensory rhodopsin
14	537.5	22.5	534	2 S55299	Htr15 transducer [
15	535.5	22.4	636	2 A84252	transducer protein
16	529	22.1	804	2 T44606	transducer protein
17	524	21.9	451	2 T44964	transducer protein
18	523.5	21.9	544	2 T44938	transducer protein
19	523.5	21.9	643	2 B84305	Htr8 transducer [i
20	521.5	21.8	628	2 F84219	Htr16 transducer [
21	521	21.8	545	2 E84327	Htr7 transducer [i
22	521	21.8	545	2 T46811	halobacterial tran
23	509.5	21.3	789	2 B84236	Htr6 transducer [i
24	507.5	21.2	788	2 T44262	transducer protein
25	506	21.1	420	2 C84298	Htr12 transducer [
26	505	21.1	642	1 T44253	transducer protein
27	502.5	21.0	773	2 T44989	transducer protein
28	499.5	20.9	627	2 F84194	Htr14 transducer [
29	497.5	20.8	419	2 T44276	transducer protein

ALIGNMENTS

RESULT 1

T44978

transducer protein hemAT [validated] - Halobacterium salinarum

N:Alternate names: methyl-accepting taxis protein htrB; transducer protein htrB; transduce

C:Species: Halobacterium salinarum

C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 15-Sep-2000

C:Accession: T44978

R:Zhang, W.; Brooun, A.; McCandless, J.; Banda, P.; Alam, M.

Proc. Natl. Acad. Sci. U.S.A. 93, 4649-4654, 1996

A:Title: Signal transduction in the archaeon Halobacterium salinarum is processed throu

A:Reference number: Z22804; MUID:96209786; PMID:8643458

A:Accession: T44978

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-489 <ZHA>

A:Cross-references: EMBL:U75436; NID:gl654420; PIDN:AAB17881.1; PID:gl654421

A:Experimental source: strain Flx15

A>Note: the source is designated as Halobacterium salinarum

C:Genetics:

A:Gene: hemAT; htp15; htrB

C:Function:

A:Description: involved in aerotactic signal transduction; involved in oxygen sensing; "

C:Superfamily: Halobacterium salinarum transducer protein htrI

C:Keywords: heme; methylated amino acid; signal transduction

Query Match 100.0%; Score 2394; DB 2; Length 489;
Best Local Similarity 100.0%; Pred. No. 7.8e-98;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSNDNDTLVTADVVRNGIDGHADRLGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT	60
DB	1	MSNDNDTLVTADVVRNGIDGHADRLGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT	60
QY	61	ADALVTDFYDHLSEYERTQDLFANSTKTVEQLKETAQYLLGLGRGEYDTYEAQARIG	120
DB	61	ADALVTDFYDHLSEYERTQDLFANSTKTVEQLKETAQYLLGLGRGEYDTYEAQARIG	120
QY	121	KIHVILGLGPDVYLGATRYTITGLLDALADDVVADRGEAAAVDELVARFLPMLKLTTF	180
DB	121	KIHVILGLGPDVYLGATRYTITGLLDALADDVVADRGEAAAVDELVARFLPMLKLTTF	180
QY	181	DDQIAMDYTDVYVAQRUHLDEIDSRQELANAVATHVEAPLSLEATSDVAERTDTMEART	240
DB	181	DDQIAMDYTDVYVAQRUHLDEIDSRQELANAVATHVEAPLSLEATSDVAERTDTMEART	240
QY	241	DDQVDRMADYSREISSVSASVEEVASTADVRRTSEDAEALAAQGEAAADALATMTDID	300
DB	241	DDQVDRMADYSREISSVSASVEEVASTADVRRTSEDAEALAAQGEAAADALATMTDID	300
QY	301	EATDGVGTAGVEQLGERAADVESVTGVDIAEQTNMLALNASIEAARAGEGFAVVAD	360
DB	301	EATDGVGTAGVEQLGERAADVESVTGVDIAEQTNMLALNASIEAARAGEGFAVVAD	360

30	497	20.8	536	2	D84325	Htr17 transducer [
31	475.5	19.9	537	2	C87302	methyl-accepting c
32	464	19.4	764	2	A84328	Htr2 transducer [i
33	463.5	19.4	536	1	A47190	transducer protein
34	463.5	19.4	536	2	E84318	Htr1 transducer [i
35	461	19.3	765	1	T44946	transducer protein
36	452.5	18.9	423	2	H84257	Htr13 transducer [
37	439.5	18.4	423	2	T44258	transducer protein
38	419.5	17.5	559	2	G84132	methyl-accepting c
39	415	17.3	555	2	D87536	methyl-accepting c
40	413	17.3	439	2	A83713	methyl-accepting c
41	412.5	17.2	432	2	C69832	chemotactic transd
42	409.5	17.1	632	2	H83106	methyl-accepting c
43	408	17.0	499	2	A97485	methyl-accepting c
44	408	17.0	499	2	AB2703	methyl-accepting c
45	405.5	16.9	579	2	D84137	methyl-accepting c

QY 361 EVKALAEBSRQSTRVEELVQMQAETETVDQLDEVNQRIGEGVERVEEAMETIQEITD 420
Db |||||
QY 361 EVKALAEBSRQSTRVEELVQMQAETETVDQLDEVNQRIGEGVERVEEAMETIQEITD 420
Db |||||
QY 421 AVEAASGMQVSTATDEQAVSTVEAEMVGVDDRAGEIAAALDDIADATDQQVTVVEE 480
Db |||||
QY 421 AVEAASGMQVSTATDEQAVSTVEAEMVGVDDRAGEIAAALDDIADATDQQVTVVEE 480
Db |||||
QY 481 VRETGVKLS 489
Db |||||
481 VRETGVKLS 489

RESULT 2
E84304
Htr10 transducer [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: E84304
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jäbl
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: E84304
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-489 <STO>
A:Cross-references: GB:AE004437; NID:g10580997; PIDN:AAG19801.1; GSPDB:GN00138
A:Gene: htr10
C:Superfamily: Halobacterium salinarum transducer protein htr1

Query Match 98.6%; Score 2360; DB 2; Length 489;
Best Local Similarity 99.0%; Pred. No. 2.4e-96;
Matches 484; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSNDNDTLTVADVRNGIDGHADRLGDEAEIATWRLSFTGIDDDTMAALAAEQPLFEAT 60
Db |||||
QY 1 MSNDNDTLTVADVRNGIDGHADRLGDEAEIATWRLSFTGIDDDTMAALAAEQPLFEAT 60
Db |||||
QY 61 ADALVTDFDHLSEYERTODLFANSTKTVEQLKETAQAEYLLGLGRGEYDTEYAAQARIG 120
Db |||||
QY 61 ADALVTDFDHLSEYERTODLFANSTKTVEQLKETAQAEYLLGLGRGEYDTEYAAQARIG 120
Db |||||
QY 121 KTHDVLGLGPDVYLGAITYTYTGLDALADDVVDGRGEAAAVDELVARFLPMLKLTFF 180
Db |||||
QY 121 KTHDVLGLGPDVYLGAITYTYTGLDALADDVVDGRGEAAAVDELVARFLPMLKLTFF 180
Db |||||
QY 181 DQOIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSDQVAERTDTMRART 240
Db |||||
QY 181 DQOIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSDQVAERTDTMRART 240
Db |||||
QY 241 DDQVDRMADVREISSVSASVEEASTADDVVRTSDEAALAAQOGEAAADALATMTDID 300
Db |||||
QY 241 DDQVDRMADVREISSVSASVEEASTADDVVRTSDEAALAAQOGEAAADALATMTDID 300
Db |||||
QY 301 EATDGTAGVQLGERAADVESVTGIDDIABQTNMLNALSIEAARAGEGEGFAVVAD 360
Db |||||
QY 301 EATDGTAGVQLGERAADVESVTGIDDIABQTNMLNALSIEAARAGEGEGFAVVAD 360
Db |||||
QY 361 EVKALAEBSRQSTRVEELVQMQAETETVDQLDEVNQRIGEGVERVEEAMETIQEITD 420
Db |||||
QY 361 EVKALAEBSRQSTRVEELVQMQAETETVDQLDEVNQRIGEGVERVEEAMETIQEITD 420
Db |||||
QY 421 AVEAASGMQVSTATDEQAVSTVEAEMVGVDDRAGEIAAALDDIADATDQQVTVVEE 480
Db |||||
QY 421 AVEAASGMQVSTATDEQAVSTVEAEMVGVDDRAGEIAAALDDIADATDQQVTVVEE 480
Db |||||
QY 481 VRETGVKLS 489

Db |||||
481 VRETGVKLS 489

RESULT 3
A84294
Htr9 transducer [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 17-May-2002
C:Accession: A84294
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jäbl
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: A84294
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-481 <STO>
A:Cross-references: GB:AE004437; NID:g10580901; PIDN:AAG19717.1; GSPDB:GN00138
A:Gene: htr9
C:Superfamily: Halobacterium salinarum transducer protein car

Query Match 26.9%; Score 643; DB 2; Length 481;
Best Local Similarity 37.6%; Pred. No. 2.1e-21;
Matches 182; Conservative 72; Mismatches 196; Indels 34; Gaps 10;

QY 15 NGIDGHADRLGDEAEIATWRLSFTGIDDDTMAALAAEQPLFEATADAL-VTFDYLHLE 73
Db |||||
QY 15 NGIDGHADRLGDEAEIATWRLSFTGIDDDTMAALAAEQPLFEATADAL-VTFDYLHLE 73
Db |||||
QY 19 NTLDVPAFA--VDQAGVAVM-----DDQIAAL-----LETAPEDAIGVTDIGERLN 63
Db |||||
QY 74 SYERTODLFANSTKTVEQLKETAQAEYLLGLGRGEYDTEYAAQARIGKTHDVLGLGPDVY 133
Db |||||
QY 74 SYERTODLFANSTKTVEQLKETAQAEYLLGLGRGEYDTEYAAQARIGKTHDVLGLGPDVY 133
Db |||||
QY 64 D---DGSRALANKVADTPIAHHEY---DGVGLADESYALLTGDYVYEDTTVAGNTDLW 116
Db |||||
QY 134 LGAYTRYTGLDALADDVVDGRGEAAAVDELVARFLPMLKLL-----TFDQOI--- 184
Db |||||
QY 117 FIATPVYHTGEPFGVI-EIVQDR-SSSARYQSELQALFGLVDTLTDAYDAGRFDAVTDIA 174
Db |||||
QY 185 AMDTVIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSDQVAERTDTMRARTDQV 244
Db |||||
QY 175 AEDTLDDVEYIQIGRNLTETFGDTLAAHITTEVNDVERLEAASQAVSESSAEIDELSTAQS 234
Db |||||
QY 245 DRMDVDSREISSVSASVEEASTADDVVRTSDEAALAAQOGEAAADALATMTDIDEATD 304
Db |||||
QY 235 TNVSTVATEVETLSATVOEIASTADEVVDTSATERLADDGSAASDADMMADVATAAD 294
Db |||||
QY 305 GVTAGVQLGERAADVESVTGIDDIABQTNMLNALSIEAARAGEGEGFAVVADDEVKA 364
Db |||||
QY 295 SVTSDVEALQNRIDEDWDVITGIAEQTNMLNALSIEAARAGEGEGFAVVADDEVKA 354
Db |||||
QY 365 LAERESQSTRVEELVQMQAETETVDQLDEVNQRIGEGVERVEEAMETIQEITDAVED 424
Db |||||
QY 355 LAEPQASWAGHIESLVSEIQITADTDTLTTTDDIEDAVAQVEDAMASFEETIATVEA 414
Db |||||
QY 425 AASGMQVSTATDEQAVSTVEAEMVGVDDRAGEIAAALDDIADATDQQVTVVEEVEET 484
Db |||||
QY 415 TAEGIEQVSDATNQASAEIEAAMVDETADLADDDITTVADIVISQTEAQSAMLHDDDES 474
Db |||||
QY 485 VGKL 488
Db |||||
475 VSEL 478

RESULT 4

T44973

transducer protein hcpIII [similarity] - Halobacterium salinarum

N;Alternate names: methyl-accepting taxis protein hta; transducer protein hta; transducer

C:Species: Halobacterium salinarum

C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-Mar-2001

C;Accession: T44973
R;Zhang, W.; Brooun, A.; McCandless, J.; Banda, P.; Alam, M.
Proc. Natl. Acad. Sci. U.S.A. 93, 4649-4654, 1996
A;Title: Signal transduction in the archaeon Halobacterium salinarum is processed through
A;Reference number: 222804; MUID:96209786; PMID:8643458
A;Accession: T44973
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-482 <ZHA>
A;Cross-references: EMBL:U75435; NID:g1654418; PIDN:AAB17880.1; PID:g1654419
A;Experimental source: strain Flx15
A;Note: the source is designated as Halobacterium salinarum
C;Genetics:
A;Gene: htr3; hta
C;Superfamily: Halobacterium salinarum transducer protein car
C;Keywords: methylated amino acid; signal transduction
F:219-472/Region: MCP signalling domain similarity

Query Match 25.5%; Score 609.5; DB 2; Length 482;
Best Local Similarity 37.2%; Pred. No. 6.2e-20;
Matches 181; Conservative 70; Mismatches 197; Indels 39; Gaps 12;

Qy 15 NGIDGHALADRIGLDEAIEAIRMISFTGIDDOTMAALAAEQPLFEATADAL-VTFDYHLE 73
Db 19 NTLDPVAPFA--VDADGAVVW-----DQIAAL-----LETAPDAIGVTDIGERLN 63
Qy 74 SVERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQARIGIKHIVLGLGPPVY 133
Db 64 D-----DGRALANKVADTPIDAHHEY---DGVGLADESVALLTGDVYVEDTIVAGNTDLW 116
Qy 134 LGAYTRYTGLDADVDVADRGEAAAVDELVARFLPMLKLL-----TFDQOI--- 184
Db 117 FIATPVYHTGEPGRVI-EIVQDR-SSSARYQSELOALFEGELVDTLDDAGRFDAVDIA 174
Qy 185 AMDTVIDSYAQLHDEIDSRQELANAVATHVEAPLSLSLEATSDVAERTDTWRARTDQV 244
Db 175 AEDTLDDDEYIGRNLTFFGTGLAAHITVENVDVERLEAAQVSSAEIDELSTAQS 234
Qy 245 DRMADVSREISSVSASVEVASTADVRRTSDEAALAOQGEAAADALATMTDIDEATD 304
Db 235 TNVSTVATEVETLSATVQELIASTADEVDVTSATAERLADGSAASDAADMADVATAAD 294
Qy 305 GVTAGVEQLGERAADVESVTGVIDDIAEQTNMLNALNASTAEARAGEGFAVVADEVKA 364
Db 295 SVTSDVEALQNRIEDIDEVDVITGIAEQTNMLNALNASTAEAPPGEGGFAVVAEEVKA 354
Qy 365 LAEERSEQTRVEELVEQMAETETVDQLDEVNQIRIGGVERVEEAMETLQEIITDAVE- 423
Db 355 LAEDAQSNAGHIESLVSEIQRTADTVTLVTTDRIEDAVAQVEDAMASFEEIVTAVEA 414
Qy 424 --DAAGMQEVSTATDEQAVSTEEVAEMVDGVDVDRAGEIAAALDDIADATDQQVTVVEV 481
Db 415 TAERASNM--VSDATNEQAASAEIAAMVDETADLADDITTAVADVSTQEAQSMHLHDL 472
Qy 482 RETVGLK 488
Db 473 DESVSEL 479

RESULT 5
H84336
Htr3 transducer [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 17-May-2002
R;NG, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leithausner, B.; Keller, K.; Cruz, R.; Danson, M.O.; Hough, D.W.; Maddocks, D.G.; Jablid
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: H84336

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-633 <STO>
A;Cross-references: GB:AE004437; NID:g10581300; PIDN:AAG20060.1; GSPDB:GN00138
C;Genetics:
A;Gene: htr3
C;Superfamily: Halobacterium salinarum transducer protein htrII

Query Match 24.9%; Score 597; DB 2; Length 633;
Best Local Similarity 33.8%; Pred. No. 3e-19;
Matches 150; Conservative 76; Mismatches 158; Indels 60; Gaps 7;

Qy 42 IDDDTMAALAAEQPLFEATADALVTFDYHLESYERT-----QDLFANSTKTVEQLKETQA 97
Db 213 LDELPAGFAGSLSQMHTREALITLDEAREDAEQTRKDAEEAARASERLNERLERRAA 272
Qy 98 EYLLGLGRGEYDTEYAAQARIGIKHIVLGLGPPVYLGAYTRYTGLDADDDVADRG 157
Db 273 -----EYSDDEMAAAA-----GDLTR-----RLDEVDVSEPM 299
Qy 158 EEAANADELVARFLPMLKLLTFDQIAMDITYDSYAQRLHDEIDSRQELANAVATHVEA 217
Db 300 QDIAEAFNDMMG-----DVEATL-----AQVRSIADAVD-----AASDVTST 336
Qy 218 PLSLEATSDVAERTDTMRATDDQVDRMADVRSREISSVSASVEEVASTADDVRRITSED 277
Db 337 SAAIRSNASQVSSVQDISADADQQRDLGTGDEVTSLSATVEEIAASADDAETVNO 396
Qy 278 ABALAQQGEAAADALATMTDIDEATDGTAGVQQLGERAADVESVTGVIDDIAEQTNML 337
Db 397 AATESERQELGEDAVAELEIEATADSVERVTALAEAVDAIGVTVITDIAEQTNML 456
Qy 338 ALNASTAEARAGEGFAVVADEVKALAEBSREOSRVEELVEQMAETETVDQLDEV 397
Db 457 ALNASTAEARADGSGDFAVVADEVKDLADEVKESATIEITLVDVQADVADTVADMSEL 516
Qy 398 NORIGEGVERVEEAMETLQEIITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDRA 457
Db 517 GDRVDAGSETTEAALALDDIGDQVEAANGSVQISDATDEQAASSTEVVTTMIDEVTDLS 576
Qy 458 GSIAAALDDIADATDQQVTVVEV 481
Db 577 DRTATESQVSAAEQAAVSEV 600

RESULT 6
T48840
transducer protein bast [validated] - Halobacterium salinarum
N;Alternate names: chemotaxis transducer protein bast; methyl-accepting taxis protein ht
C;Species: Halobacterium salinarum
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Sep-2000
C;Accession: T48840; T44981
R;Kokoeva, M.V.; Oesterhelt, D.
Mol. Microbiol. 35, 647-656, 2000
A;Title: Bast, a membrane-bound transducer protein for amino acid detection in Halobacte
A;Reference number: Z24542; MUID:20138366; PMID:10672186
A;Accession: T48840
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-805 <KOK>
A;Cross-references: EMBL:AJ245950; PIDN:CAB82572.1
A;Experimental source: strain S9
R;Zhang, W.; Brooun, A.; McCandless, J.; Banda, P.; Alam, M.
Proc. Natl. Acad. Sci. U.S.A. 93, 4649-4654, 1996
A;Title: Signal transduction in the archaeon Halobacterium salinarum is processed throu
A;Reference number: 222804; MUID:96209786; PMID:8643458
A;Accession: T44981
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 'M',13-128, 'R',130-133, 'S',135-200, 'MP',203-241, 'R',243-332, 'MR',335-401,403
<ZHA>
A;Cross-references: EMBL:U75437; NID:g1654422; PIDN:AAB17882.1; PID:g1654423
A;Experimental source: strain Flx15 derivative of S9

A;Note: the source is designated as Halobacterium salinarium

C;Genetics:

A;Gene: bsr; hpt10; htc

C;Function:

A;Description: involved in chemotactic signal transduction; mediates the response to five

C;Superfamily: Halobacterium salinarum transducer protein htrII

C;Keywords: methylated amino acid; signal transduction; transmembrane protein

Query Match

Best Local Similarity 24.9%; Score 597; DB 2; Length 805;

Matches 150; Conservative 76; Mismatches 158; Indels 60; Gaps 7;

Db 42 IDDDTMAALAEQPLFATDALVTDYDHLSEVET-----QDLFANSTKTVEQLKETQA 97

Db 385 LDELDLPGAFASISQMHTRLEALITDLEAREDAEQTRKDAEAREASERLNERLERRAA 444

Db 98 EYLLGLGRGYDEYAAQRARIGKIHVGLGPDVYLGATRYTGLLDALADDVADRG 157

Db 445 -----EYDEMMAAAA-----GDLTR-----FLDEVDVSEPM 471

Db 158 EEAANADELVARFLPMLKLLTFDQQTAMDTYIDSYAQLRHLDEIDSRQELANAVATHVEA 217

Db 472 QDIAEAFNDWNG-----DVEATL-----AQVRSIADAVD-----AASDVST 508

Db 218 PLSSLEATSDVAERTDTRARTDQVDRMADVSRISVSASVEEVAADVRRRTSED 277

Db 509 SAAERASQDSVESQDTSADADQQRDLGTGVDEVTLSATVEEIAASADVAETVNG 568

Db 278 ABALAQCGEAAADALATMTDIDEATDGTAGVEQLGERAADVSTGVIDDDIAEQTNML 337

Db 569 AATESRGQELGEDAVAELEIEATADSVERVTALAEVAIDGVTGVTIDIAEQTNML 628

Db 338 ALNASTIEARAGEGFAVVADEVKALAEESREQSTRVEELVEQMAQSTEEVTDQLDEV 397

Db 629 ALNASTIEARADKSGDGFVAVVADEVKALAEVKSATEIETLVDDVQADVATVADMSSE 688

Db 398 NORIGEGVERVEEAMETLQETDAVEDAASGMQEVSTATDEQAVSTEEVAEMVGVDDRA 457

Db 689 GDRVDAGSETIEAALADDIGQVEAANGSVQISDATDEQAASSTEEVVTWIDEVTDL 748

Db 458 GEIAAALDDIADATDQVQRTVEEV 481

Db 749 DRTATESQVSAAEQASVSEV 772

RESULT 7

B84238

Htr18 transducer [imported] - Halobacterium sp. NRC-1

C;Species: Halobacterium sp. NRC-1

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001

C;Accession: B84238

R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.

; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li

A;Title: Genome sequence of Halobacterium species NRC-1

A;Reference number: B84160; MUID:20504483; PMID:11016950

A;Accession: B84238

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-790 <STO>

A;Cross-references: GB:AE004437; NID:g10580382; PIDN:AAG19270.1; GSPDB:GN00138

C;Genetics:

A;Gene: htr18

C;Superfamily: Halobacterium salinarum transducer protein htrII

Query Match

Best Local Similarity 24.2%; Score 580; DB 2; Length 790;

Matches 160; Conservative 84; Mismatches 190; Indels 76; Gaps 11;

Db 1 MSNDNDPLVTADVNRGIDGHALADRIGLDEAEIAWRLSFTGIDG--DTMAALA---AEOP 55

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

```

QY      466  DIADATDQQVRTVEEV 481
      |||  :  |  :  :
      |||  :  |  :  :
Db      732  STADAAAEQASTLSDV 747

RESULT 10
T44849
transducer protein car [validated] - Halobacterium salinarum (strain S9)
N;Alternate names: arginine transducer protein Car; transducer protein htp11; transducer
C;Species: Halobacterium salinarum
A;Variety: strain S9
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-Sep-2000
C;Accession: T44849
R;Storch, K.F.; Rudolph, J.; Oesterhelt, D.
EMBO J. 18, 1146-1158, 1999
A;Title: Car: a cytoplasmic sensor responsible for arginine chemotaxis in the archaeon H
A;Reference number: Z22859; MUID:99164082; PMID:10064582
A;Accession: T44849
A;Status: preliminary; translated from GB/EMBL/DDBJ

```

A;Cross-references: EMBD:AD132321, NID:911021.51

A;Note: the source is designated as Halobacterium salinarum
C;Genetics: A;Gene: car; htpil; htrXI
C;Function:
A;Description: involved in chemotactic signal transduction; arginine sensor; mediates the
C;Superfamily: Halobacterium salinarum transducer protein car
C;Keywords: methylated amino acid; signal transduction

Query Match 23.4%; Score 560.5; DB 2; Length 452;
Best Local Similarity 31.3%; Pred. No. 7.9e-18;
Matches 159; Conservative 88; Mismatches 170; Indels 91; Gaps 13;

QY 10 TADVRNGIDGHALADRI-----GLDEREIA-----WELSTGIDDDTMAALAA 52
.: :.: |.: |.: |.: |.: |.:
Db 5 SSDMGGEATGEHLADELCEAYLGDNEDDGGDTQLRSLRRERDFWKHMEN-----QLVA 56
.: :.: |.: |.: |.: |.: |.:
QY 53 EQP--LFEATDALVTD----FYDLHESY-----ERTQDLFANSTKTVEQLKFTQAEYL 100
.: :.: |.: |.: |.: |.: |.:

```

57  EYPEGILLTAADGTVTHWNERFSDHMKWARSDALGEDASOVFS-----TABE-SETLPEAV 111
    QY 101  LCLGRGEYDTYAAQRARIKIHDLVGLGPDVYLGVATRYVTGLLDALADDVADRGEAA 160
    Db 112  VRTGDTVEEEE-----PHDPTDSLCOHGVPLRAPTG-----DVVGSFG-V 152
    QY 161  AAAVDELVAREFLPMLKLLTFDQOIAMDTYIDSYAQRULHEDIDSROELANAVATIVEAPLS 220
    Db 153  VPDISEKVKN-----QRELHD-----LHETVSSNVGHELS 182
    QY 221  SLEATSQDVAERTDTMRTARTODQVRMADVSREITSSVSASVEEVASTADDVRRTSEDAEA 280
    Db 183  ELSSEIDVGSFAETEAFAGEEIERMEGVADVSNQSATIEETASSAAEVSQSASQRAQD 242
    QY 281  LAQOGEAAADALATMTDIDBATDGTVTAQVGEQLGERAADVESVTGVIDDIAEQTNMLALN 340
    Db 243  RATEGEQTAEYTAIDRMGAVQESAERVNDTIDGLTSQADEMSEIIDAINDIAQDNMLALN 302
    QY 341  ASIEAARAGEGEGFPAVVADEVKALAEBSRQSTFEVELVEQMQAETEETVDOLDDEVNQR 400
    Db 303  ASIEAARAGEGEGFPAVVADEVKLSAEBSQBRADIEQMI VEMVTETDQTDADRGITTE 362
    QY 401  ICEGVERVEEAMETLQEIYTDVADAASGMQGVSTATDQAVSTEEVAEMVDPDVDRAGEI 460
    Db 363  IEALTAIVRETLDSLQETRKAVDEYATGKVEVAGATDDHAASTEQVAATTDEAVDKLTEL 422
    QY 461  AAALDDIDATDQQVRTVEEVRETGKL 488
    Db 423  EDRLNLISQIASEQHQRVAEIEDMVDEL 450

```

RESULT 11

F84327
 Htr5 transducer [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C:Accession: F84327
 R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablonka, J.; Kaper, J.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Oner, A.D.; Ehardt, H.; Lowe, T.M.; Li, A.; Title: Genome sequence of Halobacterium species NRC-1.
 A;Reference number: A84160; MUID:20504483; PMID:11016950
 A;Accession: F84327
 A;Status: preliminary
 A;Molecule type: DNA
 A;Cross-references: GB:AE004437; NID:g10581215; PIDN:AAG19986.1; GSPDB:GN00138
 C:Genetics:
 C:Superfamily: Halobacterium salinarum transducer protein htrII

Query Match 23.2%; Score 555; DB 2; Length 810;
 Best Local Similarity 32.1%; Pred. No. 2.7e-17;
 Matches 142; Conservative 85; Mismatches 174; Indels 42; Gaps 8;

QY 84 NSTKTVEQLKETQAEYLLGLGRGYDTEYAAQR-----ARIGKIHVDVLGLGPDVYLG 135
 DB 346 NTAAGVQSLSAAGAE-----IEAGNYVDVASSRRDEIGQLFASIGSMRDALVTQIDAEA 401
 QY 136 AVTRYTYTGLLDALAD-----DVADRGEAAAVDELVA-----RFLPMLKLL----- 178
 DB 402 AREQATEAQDAERAEADARAERADAKADAEALAAEALEAQAERYSDVMAACADGDLT 461
 QY 179 -----TFDQOITA--MDTYIDSYAQLHDEIDSRQELANAVAT---HVEAPLSSLEATS 226
 DB 462 RMPADDTTNEAMAAIAASFNEMLAQWEHTIIDI-GEFADAVATASEAEVGAADAERAS 520
 QY 227 QDVAERTDTMRARTDQVDRMADVSRISVSASVEEVASTADDVVRTSEDAEALAAQGE 286
 DB 521 GOVSRSVQETAGAADQRNMLDTVSGEMTDLNAAIEVAASADSVAEHSHQTAETARDGE 580
 QY 287 AAADDALATMTDIDEATDGTAGVEQLGERADVSVTGVIDDIABQTNMLANASIEAA 346
 DB 581 QTAEDAIERSLVSQEAIDATQVNEALDDQMAEISEIVDLISDIAEQTNMLANANIEAA 640
 QY 347 RAGEAGEFVAVDAEVKALAEBSREQSTRVELVQMQAETETVDQLDENVNQRIGEGVE 406
 DB 641 RADKSGDGFVAVDAEVKOLAEETQESAGDIERRITVEVQSQTATVAEAPAAEESMDAGID 700
 QY 407 RVEEMETLQETDAVEDAASGMQEVSTATDQAVSTEEVAVEMVDGVDVDRAGEIAAALDD 466
 DB 701 AVEEVYDAFTAVSDHADETDTCVQELISDTTDDQAASTEAEVAVSMTEEVADLSDSTAGEAQS 760
 QY 467 IADATDQVTVVEEVRETGVKLS 489
 DB 761 VSAAEAEQAASMEISDSVESLS 783

RESULT 12
 T46810
 halobacterial transducer protein IV [imported] - Halobacterium salinarum
 C:Species: Halobacterium salinarum
 C>Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 02-Sep-2000
 C:Accession: T46810
 R;Rudolph, J.; Nordmann, B.; Storch, K.F.; Gruenberg, H.; Rodewald, K.; Oesterhelt, D.
 FEBS Microbiol. Lett. 139, 161-168, 1996
 A;Title: A family of halobacterial transducer proteins.
 A;Reference number: 224094; MUID:96275896; PMID:8674984
 A;Accession: T46810
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-810 <RUD>
 A;Cross-references: EMBL:X95589; NID:g1435130; PIDN:CAA64841.1; PID:g1435131

A;Experimental source: strain S9

C;Genetics:

A;Gene: htrIV

C:Superfamily: Halobacterium salinarum transducer protein htrII

Query Match 23.1%; Score 554; DB 2; Length 810;
 Best Local Similarity 32.1%; Pred. No. 3e-17;
 Matches 142; Conservative 85; Mismatches 174; Indels 42; Gaps 8;

QY 84 NSTKTVEQLKETQAEYLLGLGRGYDTEYAAQR-----ARIGKIHVDVLGLGPDVYLG 135
 DB 346 NTAAGVQSLSAAGAE-----IEAGNYVDVASSRRDEIGQLFASIGSMRDALVTQIDAEA 401

QY 136 AVTRYTYTGLLDALAD-----DVADRGEAAAVDELVA-----RFLPMLKLL----- 178
 DB 402 AREQATEAQDAERAEADARAERADAKADAEALAAEALEAQAERYSDVMAACADGDLT 461

QY 179 -----TFDQOITA--MDTYIDSYAQLHDEIDSRQELANAVAT---HVEAPLSSLEATS 226
 DB 462 RMPADDTTNEAMAAIAASFNEMLAQWEHTIIDI-GEFADAVATASEAEVGAADAERAS 520

QY 227 QDVAERTDTMRARTDQVDRMADVSRISVSASVEEVASTADDVVRTSEDAEALAAQGE 286
 DB 521 GOVSRSVQETAGAADQRNMLDTVSGEMTDLNAAIEVAASADSVAEHSHQTAETARDGE 580

QY 287 AAADDALATMTDIDEATDGTAGVEQLGERADVSVTGVIDDIABQTNMLANASIEAA 346
 DB 581 QTAEDAIERSLVSQEAIDATQVNEALDDQMAEISEIVDLISDIAEQTNMLANANIEAA 640

QY 347 RAGEAGEFVAVDAEVKALAEBSREQSTRVELVQMQAETETVDQLDENVNQRIGEGVE 406
 DB 641 RADKSGDGFVAVDAEVKOLAEETQESAGDIERRITVEVQSQTATVAEAPAAEESMDAGID 700

QY 407 RVEEMETLQETDAVEDAASGMQEVSTATDQAVSTEEVAVEMVDGVDVDRAGEIAAALDD 466
 DB 701 AVEEVYDAFTAVSDHADETDTCVQELISDTTDDQAASTEAEVAVSMTEEVADLSDSTAGEAQS 760

QY 467 IADATDQVTVVEEVRETGVKLS 489
 DB 761 VSAAEAEQAASMEISDSVESLS 783

RESULT 13

T44597

transducer protein htrVI [similarity] - Halobacterium salinarum

N;Alternate names: methyl-accepting transducer protein htrD; transducer protein htrp

C:Species: Halobacterium salinarum

C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-Sep-2000

C:Accession: T44597

R;Zhang, W.; Brooun, A.; McCandless, J.; Banda, P.; Alan, M.

Proc. Natl. Acad. Sci. U.S.A. 93, 4649-4654, 1996

A;Title: Signal transduction in the archaeon Halobacterium salinarum is processed through

A;Reference number: 222804; MUID:96209786; PMID:8643458

A;Accession: T44597

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-777 <ZHA>

A;Cross-references: EMBL:U75438; NID:g1654424; PIDN:AAB17883.1; PID:g1654425

A;Experimental source: strain mFLX15

A;Note: the source is designated as Halobacterium salinarum

C;Genetics:

A;Gene: htrVI; htrp6; htrD

C:Superfamily: Halobacterium salinarum transducer protein htrII

C;Keywords: methylated amino acid; signal transduction; transmembrane protein

F;494-746/Region: MCP signalling domain similarity

Query Match 22.6%; Score 540.5; DB 2; Length 777;

Best Local Similarity 32.4%; Pred. No. 1.1e-16;

Matches 141; Conservative 74; Mismatches 171; Indels 49; Gaps 6;

QY 88 TVEQLKETQAEYLLGLGRGYDTEYAAQRA-RIGKIHVDVLGLGPDVYLGATRYTGTGLD 146
 DB 320 TVRALNDLEAK-APALERGEYDIDLQVAVDELGRLEAFASLRDAVQRAQSDANEQQVD 378

147 ALA-----DDVADRGEERAAAADVELVARFLP 173
Db 379 AEAARSEARAAQPEAAQAEAAEESDAGAPGDDRRGVLGDDACVAAGDLTVR--- 435
174 MLKLLTFDQQA---MDTYIDSYAQRHLDEIDSRQELANAVA---THVEAPLSLEATS 226
Qy 436 ---LDADVEQAAMADMERAFNEFMAADMEATIAERGAFADEVATASTDASAAAVEQTG 491
Db 227 ODVAERTMTARTDQVDRMADYSREISSVSASVEEASTADDVRRRTSEDAEALAOQGE 286
Qy 492 RDVSTAVGRIRDRADQDQLEAVASETDMSATIEEVAVAGQVAETSQRAAALGDDGQ 551
Db 287 AAADDALATMTDIDEATDGVTAGVQOLGERAADVESVTGVIDDIAEQTNMLNALSIAEA 346
Qy 552 AAAQDAVAQLEIEDETQAAATAVDDLEAKNSEIETIVAAITDIAEQTNMLNALSIAEA 611
Db 347 RAGEAGEFVAVADVVKALAEBSREQSTVEBELVQMQAETEETVDQLDENVNORIGEGVE 406
Qy 612 RADQDGDGFVAVADVVKALADESKALAEPMQALVAEVRQAQTSVAAAMDRIQERVSDGVE 671
Db 407 RVEEAMETLOETTDADVEDAAQMGQVSTATDRQAVSTVEEAVEMVDGVDVDRAGEIAAALDD 466
Qy 672 TVSETERSLSDTAGRIAEPTDGVQELSNAMDQAAASVSDVTTAVGDVAALGEETATEAES 731
Db 467 IADATDQQVTVVEV 481
Qy 732 TRDAAAEQATTLSDV 746
Db

RESULT 14
S55299
sensory rhodopsin II transducer protein - Natronobacterium pharaonis
C:Species: Natronobacterium pharaonis
C:Date: 14-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 02-Sep-2000
C:Accession: S55299
R:Seidel, R.; Scharf, B.; Gautel, M.; Kleine, K.; Oesterhelt, D.; Engelhard, M.
Proc. Natl. Acad. Sci. U.S.A. 92, 3036-3040, 1995
A:Title: The primary structure of sensory rhodopsin II: a member of an additional retinal
A:Reference number: S55296; MUID:95224074; PMID:7708770
A:Accession: S55299
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-534 <SEI>
A:Cross-references: EMBL:Z35086; NID:g510868; PID:g510869
A:Experimental source: strain SP1
A:Note: it is uncertain whether Met-1 or Met-22 is the initiator
C:Genetics:

A:Gene: htrII
C:Superfamily: Halobacterium salinarum transducer protein htrI
C:Keywords: signal transduction; transmembrane protein
F:22-42/Domain: transmembrane #status predicted <TM1>
F:59-79/Domain: transmembrane #status predicted <TM2>

Query Match 22.5%; Score 537.5; DB 2; Length 534;
Best Local Similarity 31.2%; Pred. No. 9.7e-17;
Matches 150; Conservative 95; Mismatches 177; Indels 59; Gaps 12;

Qy 16 GIDGHALADRLGLDEA---EIAWRLSFTGIDDDTMAALAAEQPLFEATADALVDFY-- 69
Db 72 GINLGLVAATLGGTGAASLSTLAAKASRMG-DGLDLVEL-----ETRDEIDGLYAA 123
Qy 70 -DHLESVERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARTGKTHDVGL 128
Db 124 FDEMRSQVRSLSLEDAKNAREDAEQAKRAE-----EINTELQAEAREFGEVMDRCAD 175
Qy 129 GPDVYLGAYTRYTGLLDALADDVVDVDRGEBAARAADELVARFLPMLKLLTFDQQAAMD 188
Db 176 GD-----FTQRLDAETDN-----FAMQSIEG-----SFNEW--MDG 204
Qy 189 YIDSYAQRHLDEIDSRQELANAVATHVEAPLSLSLEATSQDVAERTDTMTARTDQVDRMA 248
Db 205 -IEALVGRIERFADAVSDAEAVRANAE---SVMEA-SEDEVNRAVQNSIDRAAGDQQTETVQ 259

249 DVSREISSVSASVEEASTADDVRRRTSEDAEALAOQGEAADDALATMTDIDEATDGVTA 308
Db 260 QIALEUMDDVSTAEVVAASADDIAKTARQAETGEAGRETAETAITEMNEVESRTEQAVA 319
Qy 309 GVEOLGERAADVESVTGVIDDIAEQTNMLNALSIAERAGEAGEGFVAVADEVKALAE 368
Db 320 SMEELNEDEVREIGEVSSEMIADIAEQTNLALNASIARADNGSEGFAVVADEVKALAE 379
Qy 369 SREOSTRVEELVQMQAETEETVDQLDENVNORIGEGVERVEEAMETLOETTDADVEDAA 428
Db 380 TKAATEEIDDLIGTVQDRTQTTVDIIRTSQVSEGVETVDBLERIVDSVVRTNDG 439
Qy 429 MOEVSTATDEQAVSTEEVAEMVDGVDVDRAGISTAAALDDIADATDQVTVVEVRETGKL 488
Db 440 IQEINQSIDAQAQAQAATTWVEDMAATSEQTASDAETAETTTQAEISVKEVFOLLIDL 499
Qy 489 S 489
Db 500 S 500

RESULT 15
A84252
Htr15 transducer [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84252
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: A84252
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-636 <STO>
A:Cross-references: GB:AE004437; NID:g10580513; PIDN:AAG19381.1; GSPDB:GN00138
C:Genetics:
A:Gene: htr15

Query Match 22.4%; Score 535.5; DB 2; Length 636;
Best Local Similarity 29.7%; Pred. No. 1.5e-16;
Matches 149; Conservative 99; Mismatches 191; Indels 63; Gaps 11;

Qy 17 IDGHALADRLGLDEAETIAWRLSFTGIDDDTMAALAAEQPLFEAT---ADALVTD----- 67
Db 165 VDDDALDLSIGMP---VFMLDATG-----SVVAVNETIEELTGCSKADAMGMHASEA 214
Qy 68 FYDHLSEYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDLVG 127
Db 215 FYDPRRVKTLAD-----KVIESPRSA--EFDIERSEKAQLYADTSMV- 258
Qy 128 LGPDVYLGAYTRYVYTGILLDALDD-----VVADRGEA--AAAVDELVARFLPML- 175
Db 259 ----TDGGTDRHIRFKFASPIFDDGCELLAAVEATIEDTVRRADAVEBLVDLSTTID 314
Qy 176 -----KLLTFDQQAAMDYIDSYAQRHLDEIDSRQELANAVATHVEAPLSLEATSQ 227
Db 315 ALSSGQLSKRASFEHIGIINEQLVSVVVSALNGMADQFERLVGVQVDGQTQLADTIERATA 374
Qy 228 DVAERTTMTARTDQVDRMADYSREISSVSASVEEASTADDVRRRTSEDAEALAOQGEA 287
Db 375 DATDIADTV-----SSQNEMLSSAANEMENFSAQMGEVAASDDQVSAARQAQDAESGLE 430
Qy 288 AADDALATMTDIDEATDGVTAGVQOLGERAADVESVTGVIDDIAEQTNMLNALSIAEAA 347
Db 431 ASEGANQATNEVIDISDDLMEVSKLESRMDEIEDVVEVIAEAVADQTNLLNALSIAEAA 490
Qy 348 AGEAGEGFVAVADVVKALAEBSREQSTVEBELVQMQAETEETVDQLDENVNORIGEGVE 407

Db	491	AGEAGSGFAVVAVEKELANETREHTERIAGSISDVQQQANETVLAVEESHEQIHRAGDE	550
Qy	408	VEEAMETLQEIITDAVEDAASGMQEVSTATDEQAVSTEEVVAEWMVGVDDRAGEIAAALDDI	467
Db	551	IDDALTALEEIATSVDAAATGITEVARANDEQASTVEDVITIEDVQQQAEAAAAASDRI	610
Qy	468	ADATDQQRVVEEVRETVGKLS	489
Db	611	VSATQEQSTAVSQLSERVDKLT	632

Search completed: August 10, 2004, 15:29:26
Job time : 22.3123 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:19:00 ; Search time 15.323 Seconds
(without alignments)
1654.229 Million cell updates/sec

Title: US-09-455-978B-2
Perfect score: 2394
Sequence: 1 MSNDNDTLVTADVRNGIDGH.....ATDQQVTVBEVRETVGKLS 489

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2394	100.0	489	1 HMAT_HALSA	P71413 halobacteri
2	2360	98.6	489	1 HMAT_HALN1	Q9hpr6 halobacteri
3	643	26.9	481	1 HTR3_HALN1	Q9hq00 halobacteri
4	643	26.9	481	1 HTR3_HALSA	Q48316 halobacteri
5	578	24.1	778	1 HTR6_HALN1	Q9hr92 halobacteri
6	573	23.9	778	1 HTR6_HALSA	Q48319 halobacteri
7	555	23.2	810	1 HTR4_HALN1	Q9hp84 halobacteri
8	554	23.1	810	1 HTR4_HALSA	Q48317 halobacteri
9	537.5	22.5	534	1 HTR2_NATPH	P42259 natronomona
10	521	21.8	545	1 HTR5_HALN1	Q48318 halobacteri
11	492	20.6	433	1 HTR2_HALVA	P42258 haloarcula
12	464	19.4	763	1 HTR2_HALN1	Q9hp81 halobacteri
13	463.5	19.4	535	1 HTR1_HALSA	P33741 halobacteri
14	461	19.3	764	1 HTR2_HALSA	P71410 halobacteri
15	449.5	18.8	535	1 HTR1_HALSA	P33955 halobacteri
16	412.5	17.2	432	1 HMAT_BACSU	O07621 bacillus su
17	388.5	16.2	662	1 MCPB_BACSU	P39215 bacillus su
18	386.5	16.1	682	1 PILJ_PSEAF	P42257 pseudomonas
19	379	15.8	662	1 TLPB_BACSU	P39217 bacillus su
20	359.5	15.0	547	1 MCPFC_SALTY	Q02755 salmonella
21	354.5	15.0	1000	1 Y041_SYNY3	Q55445 synechocyst
22	354.5	14.8	662	1 TLPA_BACSU	P159216 bacillus chol
23	351	14.7	548	1 HLYB_VIBCH	P159217 vibrio chol
24	349	14.6	553	1 MCP2_ECOLI	P02942 escherichia
25	348.5	14.6	551	1 MCP1_ECOLI	P02942 escherichia
26	347.5	14.5	573	1 TLPC_BACSU	P39209 bacillus su
27	342.5	14.3	657	1 MCPA_CAUCR	Q00986 caulobacter
28	335.5	14.0	553	1 MCP2_SALTY	P02941 salmonella
29	334	14.0	661	1 MCPA_BACSU	P39214 bacillus su
30	333.5	13.9	654	1 MCP3_BACSU	P54576 bacillus su
31	326.5	13.6	546	1 MCP3_ECOLI	P05704 escherichia
32	323.5	13.5	668	1 DCRA_DESVH	P35841 desulfovibr
33	318.5	13.3	557	1 MCP5_ENTAE	P21822 enterobacte

RESULT 1
HMAT_HALSA
ID HMAT_HALSA STANDARD; PRT; 489 AA.
AC P71413.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heme-based aerotactic transducer hemar (Transducer HtB protein).
GN HEMAT OR HTB.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PLX15.
RX MEDLINE=96209786; PubMed=8643458;
RA Zhang W., Brooun A., McCandless J., Banda P., Alam M.;
RT "Signal transduction in the archaeon Halobacterium salinarum is
processed through three subfamilies of 13 soluble and membrane-bound
transducer proteins";
RL Proc. Natl. Acad. Sci. U.S.A. 93:4649-4654(1996).
RN [2]
RP FUNCTION.
MEDLINE=20140131; PubMed=10676961;
RA Hou S., Larsen R.W., Boudko D., Riley C.W., Karatan E., Zimmer M.,
Ordal G.W., Alam M.;
RT "Myoglobin-like aerotaxis transducers in Archaea and Bacteria";
RL Nature 403:540-544(2000).
CC -|- FUNCTION: Heme-containing signal transducer responsible for
aerotaxis, the migratory response toward or away from oxygen.
CC -|- SUBUNIT: Homotrimer (Probable).
CC -|- SIMILARITY: Contains 1 methyl-accepting transducer domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

CC EMBL; U75436; AAB17881.1; -;
DR PIR; T44978; T44978.
DR HSSP; P02942; 10U7.
DR GO; GO:0020037; F:heme binding; IDA.
DR GO; GO:0004871; F:signal transducer activity; IDA.
DR GO; GO:0009454; P:aerotaxis; IDA.
DR GO; GO:0007165; P:signal transduction; IDA.
DR InterPro; IPR004089; Chmtaxis trans.
DR InterPro; IPR004090; Me chemotaxis.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS0111; CHEMOTAXIS_TRANSDUC_2; 1.
KW Transducer; Heme.

34 318 13.3 533 1 MCPE RHIME
35 313.5 13.1 557 1 CPS CLOTH
36 311.5 13.0 512 1 MCPD ENTAE
37 309 12.9 756 1 Y4SI_RHISN
38 308 12.9 506 1 AER_ECOLI
39 304 12.7 533 1 MCP4_ECOLI
40 299.5 12.5 845 1 Y4FA_RHISN
41 286 11.9 620 1 TCPI_VIBCH
42 271.5 11.3 417 1 FRZC MYXXA
43 202.5 8.5 955 1 KINL_LEICH
44 196.5 8.2 1972 1 MYHB_HUMAN
45 195.5 8.2 1972 1 MYHB_RABIT

Q52877 rhizobium m
Q02929 clostridium
P21823 enterobacte
P55652 rhizobium s
P50466 escherichia
P07018 escherichia
P55439 rhizobium s
P29486 vibrio chol
P43500 myxococcus
P46865 leishmania
P35749 homo sapien
P35748 oryctolagus

```

FT DOMAIN 218 454 METHYL-ACCEPTING TRANSDUCER.
SQ SEQUENCE 489 AA; 52816 MW; 8457263FCF616BFF CRC64;

Query Match
Best Local Similarity 100.0%; Score 2394; DB 1; Length 489;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNDNDTLVADVNRNGIDGHADRIGLDEAIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
Db 1 MSNDNDTLVADVNRNGIDGHADRIGLDEAIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
QY 61 ADALVTFDYHLESYERTQDLFANSTKTVEQLKETAQAEVLLGLGRGEYDTYAAQARIG 120
Db 61 ADALVTFDYHLESYERTQDLFANSTKTVEQLKETAQAEVLLGLGRGEYDTYAAQARIG 120
QY 121 KIHVDVLGDPVYLGAYTRYTGLLDALADDVADRGEEAAAADDELVARFLPMLKLTFF 180
Db 121 KIHVDVLGDPVYLGAYTRYTGLLDALADDVADRGEEAAAADDELVARFLPMLKLTFF 180
QY 181 DQIAMDYIDSYAORLHDEIDSRQELANAVATHVEAPLSLEATSDVAERTDTMRAAT 240
Db 181 DQIAMDYIDSYAORLHDEIDSRQELANAVATHVEAPLSLEATSDVAERTDTMRAAT 240
QY 241 DDQVDMADVSREISSVSASVEEASTADVVRTSDEAALAAQGEAAADALATMTDID 300
Db 241 DDQVDMADVSREISSVSASVEEASTADVVRTSDEAALAAQGEAAADALATMTDID 300
QY 301 EATDGTAGVQELGERAADVESVTGIDDIAGQTNMLNALSIEAARAGEGFAVVAD 360
Db 301 EATDGTAGVQELGERAADVESVTGIDDIAGQTNMLNALSIEAARAGEGFAVVAD 360
QY 361 EVKALAEBSREQSTRVEELVEQMAETETVDQLDEVNQRIGGVERVEEAMETLQETID 420
Db 361 EVKALAEBSREQSTRVEELVEQMAETETVDQLDEVNQRIGGVERVEEAMETLQETID 420
QY 421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDVDRAGEIAAALDDIADATDQOVRTVEE 480
Db 421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDVDRAGEIAAALDDIADATDQOVRTVEE 480
QY 481 VRETGVKLS 489
Db 481 VRETGVKLS 489

RESULT 2
ID HMAT HALN1 STANDARD; PRT; 489 AA.
AC Q5HPR5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Heme-based aerotactic transducer hemAT.
GN HMAT OR HTR10 OR VNG1505G.
OC Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahaitas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonksi P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Ikenbarger T.A., Peck R.F., Pohlischer M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -!- FUNCTION: Heme-containing signal transducer responsible for
CC aerotaxis, the migratory response toward or away from oxygen (By
CC similarity).

```

```

CC -!- SUBUNIT: Homotetramer (Probable).
CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; AE005064; AAG19801.1; --
CC PIR; E84304; E84304.
CC HSSP; P02942; 10U7.
CC InterPro; IPR004089; Chmtaxis transd.
CC Pfam; PF00015; MCPsinal; 1.
CC PRINTS; PR00260; CHEMTNSDUCR.
CC SMART; SM00283; MA; 1.
CC PROSITE; PS00111; CHEMOTAXIS_TRANSDUC_2; 1.
CC Transducer; Heme; Complete proteome.
KW DOMAIN 218 454 METHYL-ACCEPTING TRANSDUCER.
SQ SEQUENCE 489 AA; 52852 MW; 8C0AEF179667791E CRC64;

Query Match
Best Local Similarity 99.0%; Score 2360; DB 1; Length 489;
Matches 484; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSNDNDTLVADVNRNGIDGHADRIGLDEAIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
Db 1 MSNDNDTLVADVNRNGIDGHADRIGLDEAIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
QY 61 ADALVTFDYHLESYERTQDLFANSTKTVEQLKETAQAEVLLGLGRGEYDTYAAQARIG 120
Db 61 ADALVTFDYHLESYERTQDLFANSTKTVEQLKETAQAEVLLGLGRGEYDTYAAQARIG 120
QY 121 KIHVDVLGDPVYLGAYTRYTGLLDALADDVADRGEEAAAADDELVARFLPMLKLTFF 180
Db 121 KIHVDVLGDPVYLGAYTRYTGLLDALADDVADRGEEAAAADDELVARFLPMLKLTFF 180
QY 181 DQIAMDYIDSYAORLHDEIDSRQELANAVATHVEAPLSLEATSDVAERTDTMRAAT 240
Db 181 DQIAMDYIDSYAORLHDEIDSRQELANAVATHVEAPLSLEATSDVAERTDTMRAAT 240
QY 241 DDQVDMADVSREISSVSASVEEASTADVVRTSDEAALAAQGEAAADALATMTDID 300
Db 241 DDQVDMADVSREISSVSASVEEASTADVVRTSDEAALAAQGEAAADALATMTDID 300
QY 301 EATDGTAGVQELGERAADVESVTGIDDIAGQTNMLNALSIEAARAGEGFAVVAD 360
Db 301 EATDGTAGVQELGERAADVESVTGIDDIAGQTNMLNALSIEAARAGEGFAVVAD 360
QY 361 EVKALAEBSREQSTRVEELVEQMAETETVDQLDEVNQRIGGVERVEEAMETLQETID 420
Db 361 EVKALAEBSREQSTRVEELVEQMAETETVDQLDEVNQRIGGVERVEEAMETLQETID 420
QY 421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDVDRAGEIAAALDDIADATDQOVRTVEE 480
Db 421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDVDRAGEIAAALDDIADATDQOVRTVEE 480
QY 481 VRETGVKLS 489
Db 481 VRETGVKLS 489

RESULT 3
ID HTR3 HALN1 STANDARD; PRT; 481 AA.
AC Q9H000;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Halobacterial transducer protein III.

```

HTR9 OR VNG1395G.
 Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
 Archaea; Buryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 NCBI_TaxID=64091;
 [1]_SEQUENCE FROM N.A.
 MEDLINE=20504483; PubMed=11016950;
 NG W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
 RT "Genome sequence of Halobacterium species NRC-1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).
 CC -!- FUNCTION: Potentially involved in chemo- or phototactic signal
 CC transduction (by similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
 CC -!- SIMILARITY: Contains 1 PAS (PER-ARNT-SIM) dimerization domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE005058; AAG19717.1; -.
 DR PIR; A84294; A84294.
 DR HSSP; P02942; 1Q07.
 DR InterPro; IPR004089; Chmtaxis_transd.
 DR InterPro; IPR003660; HAMP.
 DR InterPro; IPR004090; Me_chemotaxis.
 DR InterPro; IPR000014; PAS_domain.
 DR Pfam; PF00015; MCPsignal; 1.
 DR PRINTS; PR00260; CHEMTRNSDUCR.
 DR SMART; SM00304; HAMP; 1.
 DR SMART; SM00283; PAS; 1.
 DR SMART; SM00091; PAS; 1.
 DR PROSITE; PS01111; CHEMOTAXIS_TRANSDUC_2; 1.
 DR PROSITE; PS01112; PAS; 1.
 DR TRANSFAC; TF00001; PAS; 1.
 DR TRANSFAC; TF00002; PAS; 1.
 DR TRANSFAC; TF00003; PAS; 1.
 DR TRANSFAC; TF00004; PAS; 1.
 DR TRANSFAC; TF00005; PAS; 1.
 DR TRANSFAC; TF00006; PAS; 1.
 DR TRANSFAC; TF00007; PAS; 1.
 DR TRANSFAC; TF00008; PAS; 1.
 DR TRANSFAC; TF00009; PAS; 1.
 DR TRANSFAC; TF00010; PAS; 1.
 DR TRANSFAC; TF00011; PAS; 1.
 DR TRANSFAC; TF00012; PAS; 1.
 DR TRANSFAC; TF00013; PAS; 1.
 DR TRANSFAC; TF00014; PAS; 1.
 DR TRANSFAC; TF00015; PAS; 1.
 DR TRANSFAC; TF00016; PAS; 1.
 DR TRANSFAC; TF00017; PAS; 1.
 DR TRANSFAC; TF00018; PAS; 1.
 DR TRANSFAC; TF00019; PAS; 1.
 DR TRANSFAC; TF00020; PAS; 1.
 DR TRANSFAC; TF00021; PAS; 1.
 DR TRANSFAC; TF00022; PAS; 1.
 DR TRANSFAC; TF00023; PAS; 1.
 DR TRANSFAC; TF00024; PAS; 1.
 DR TRANSFAC; TF00025; PAS; 1.
 DR TRANSFAC; TF00026; PAS; 1.
 DR TRANSFAC; TF00027; PAS; 1.
 DR TRANSFAC; TF00028; PAS; 1.
 DR TRANSFAC; TF00029; PAS; 1.
 DR TRANSFAC; TF00030; PAS; 1.
 DR TRANSFAC; TF00031; PAS; 1.
 DR TRANSFAC; TF00032; PAS; 1.
 DR TRANSFAC; TF00033; PAS; 1.
 DR TRANSFAC; TF00034; PAS; 1.
 DR TRANSFAC; TF00035; PAS; 1.
 DR TRANSFAC; TF00036; PAS; 1.
 DR TRANSFAC; TF00037; PAS; 1.
 DR TRANSFAC; TF00038; PAS; 1.
 DR TRANSFAC; TF00039; PAS; 1.
 DR TRANSFAC; TF00040; PAS; 1.
 DR TRANSFAC; TF00041; PAS; 1.
 DR TRANSFAC; TF00042; PAS; 1.
 DR TRANSFAC; TF00043; PAS; 1.
 DR TRANSFAC; TF00044; PAS; 1.
 DR TRANSFAC; TF00045; PAS; 1.
 DR TRANSFAC; TF00046; PAS; 1.
 DR TRANSFAC; TF00047; PAS; 1.
 DR TRANSFAC; TF00048; PAS; 1.
 DR TRANSFAC; TF00049; PAS; 1.
 DR TRANSFAC; TF00050; PAS; 1.
 DR TRANSFAC; TF00051; PAS; 1.
 DR TRANSFAC; TF00052; PAS; 1.
 DR TRANSFAC; TF00053; PAS; 1.
 DR TRANSFAC; TF00054; PAS; 1.
 DR TRANSFAC; TF00055; PAS; 1.
 DR TRANSFAC; TF00056; PAS; 1.
 DR TRANSFAC; TF00057; PAS; 1.
 DR TRANSFAC; TF00058; PAS; 1.
 DR TRANSFAC; TF00059; PAS; 1.
 DR TRANSFAC; TF00060; PAS; 1.
 DR TRANSFAC; TF00061; PAS; 1.
 DR TRANSFAC; TF00062; PAS; 1.
 DR TRANSFAC; TF00063; PAS; 1.
 DR TRANSFAC; TF00064; PAS; 1.
 DR TRANSFAC; TF00065; PAS; 1.
 DR TRANSFAC; TF00066; PAS; 1.
 DR TRANSFAC; TF00067; PAS; 1.
 DR TRANSFAC; TF00068; PAS; 1.
 DR TRANSFAC; TF00069; PAS; 1.
 DR TRANSFAC; TF00070; PAS; 1.
 DR TRANSFAC; TF00071; PAS; 1.
 DR TRANSFAC; TF00072; PAS; 1.
 DR TRANSFAC; TF00073; PAS; 1.
 DR TRANSFAC; TF00074; PAS; 1.
 DR TRANSFAC; TF00075; PAS; 1.
 DR TRANSFAC; TF00076; PAS; 1.
 DR TRANSFAC; TF00077; PAS; 1.
 DR TRANSFAC; TF00078; PAS; 1.
 DR TRANSFAC; TF00079; PAS; 1.
 DR TRANSFAC; TF00080; PAS; 1.
 DR TRANSFAC; TF00081; PAS; 1.
 DR TRANSFAC; TF00082; PAS; 1.
 DR TRANSFAC; TF00083; PAS; 1.
 DR TRANSFAC; TF00084; PAS; 1.
 DR TRANSFAC; TF00085; PAS; 1.
 DR TRANSFAC; TF00086; PAS; 1.
 DR TRANSFAC; TF00087; PAS; 1.
 DR TRANSFAC; TF00088; PAS; 1.
 DR TRANSFAC; TF00089; PAS; 1.
 DR TRANSFAC; TF00090; PAS; 1.
 DR TRANSFAC; TF00091; PAS; 1.
 DR TRANSFAC; TF00092; PAS; 1.
 DR TRANSFAC; TF00093; PAS; 1.
 DR TRANSFAC; TF00094; PAS; 1.
 DR TRANSFAC; TF00095; PAS; 1.
 DR TRANSFAC; TF00096; PAS; 1.
 DR TRANSFAC; TF00097; PAS; 1.
 DR TRANSFAC; TF00098; PAS; 1.
 DR TRANSFAC; TF00099; PAS; 1.
 DR TRANSFAC; TF00100; PAS; 1.
 DR TRANSFAC; TF00101; PAS; 1.
 DR TRANSFAC; TF00102; PAS; 1.
 DR TRANSFAC; TF00103; PAS; 1.
 DR TRANSFAC; TF00104; PAS; 1.
 DR TRANSFAC; TF00105; PAS; 1.
 DR TRANSFAC; TF00106; PAS; 1.
 DR TRANSFAC; TF00107; PAS; 1.
 DR TRANSFAC; TF00108; PAS; 1.
 DR TRANSFAC; TF00109; PAS; 1.
 DR TRANSFAC; TF00110; PAS; 1.
 DR TRANSFAC; TF00111; PAS; 1.
 DR TRANSFAC; TF00112; PAS; 1.
 DR TRANSFAC; TF00113; PAS; 1.
 DR TRANSFAC; TF00114; PAS; 1.
 DR TRANSFAC; TF00115; PAS; 1.
 DR TRANSFAC; TF00116; PAS; 1.
 DR TRANSFAC; TF00117; PAS; 1.
 DR TRANSFAC; TF00118; PAS; 1.
 DR TRANSFAC; TF00119; PAS; 1.
 DR TRANSFAC; TF00120; PAS; 1.
 DR TRANSFAC; TF00121; PAS; 1.
 DR TRANSFAC; TF00122; PAS; 1.
 DR TRANSFAC; TF00123; PAS; 1.
 DR TRANSFAC; TF00124; PAS; 1.
 DR TRANSFAC; TF00125; PAS; 1.
 DR TRANSFAC; TF00126; PAS; 1.
 DR TRANSFAC; TF00127; PAS; 1.
 DR TRANSFAC; TF00128; PAS; 1.
 DR TRANSFAC; TF00129; PAS; 1.
 DR TRANSFAC; TF00130; PAS; 1.
 DR TRANSFAC; TF00131; PAS; 1.
 DR TRANSFAC; TF00132; PAS; 1.
 DR TRANSFAC; TF00133; PAS; 1.
 DR TRANSFAC; TF00134; PAS; 1.
 DR TRANSFAC; TF00135; PAS; 1.
 DR TRANSFAC; TF00136; PAS; 1.
 DR TRANSFAC; TF00137; PAS; 1.
 DR TRANSFAC; TF00138; PAS; 1.
 DR TRANSFAC; TF00139; PAS; 1.
 DR TRANSFAC; TF00140; PAS; 1.
 DR TRANSFAC; TF00141; PAS; 1.
 DR TRANSFAC; TF00142; PAS; 1.
 DR TRANSFAC; TF00143; PAS; 1.
 DR TRANSFAC; TF0014


```

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Halobacterial transducer protein VI.
GN HTR4 OR HTPV1.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S9;
RX MEDLINE=96275896; PubMed=8674984;
RA Rudolf J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,
RA Oesterhelte D.;
RT "A family of halobacterial transducer proteins.";
RL FEMS Microbiol. Lett. 139:161-168 (1996).
CC -!- FUNCTION: Potentially involved in chemo- or phototactic signal
transduction.
CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC -!- SIMILARITY: Contains 2 HAMP domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DE EMBL; X95590; CAA64843.1; -.
DR PIR; T48897; T48897.
DR HSSP; P02942; 1QU7.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me_Chemotaxis.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMP; 2.
DR Transducer; Transmembrane; Repeat.
KW DOMAIN 1 26 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 27 47 POTENTIAL.
FT DOMAIN 48 296 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 297 317 POTENTIAL.
FT DOMAIN 318 778 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 318 370 HAMP 1.
FT DOMAIN 412 465 HAMP 2.
FT DOMAIN 484 720 METHYL-ACCEPTING TRANSDUCER.
SQ SEQUENCE 778 AA; 82077 MW; 134C7D7F0A3334CD CRC64;

Query Match 23.9%; Score 573; DB 1; Length 778;
Best Local Similarity 33.0%; Pred. No. 1.3e-17;
Matches 144; Conservative 74; Mismatches 168; Indels 50; Gaps 6;

QY 88 TVEQLKETQAEVLLGLGRGEYDTEYAAQRA-RIGKIHVDVLGPDVYLGAITYRYTGLLD 146
DB 320 TVRALNDLEAK-AAALERGERYDITLDVAVKVDLGRLEFAFASLRDTVQARIDANEQQVD 378
QY 147 -----ALADVDVADRGEAAAAVDLVARFL-----PMLKLLTFDQOIAMD 187
DB 379 ABAARSEAAQADAEAAQAEAAEAESEAAQARLETTAEAFSETMAYAYAGDLTVRLD 438
QY 188 TVIDSYAQ-----RUHDEIDSRQELANAVATHVEAPLSLEAT 225
DB 439 ADVEQAAMADIAAFAFNEMAAADMEATIADAGRADEVAT-----ASTDASDAAAVEOT 491
QY 226 SDQVAERTDTWEARTDDQVDRWADYSRELSVSAVEEVAASTADDVRITSEDAEALAQG 285
DB 492 GRDVSADVGRIRDAARDQRLQLEAVASETDEMSATIEEVAASDAQVETSQRAAALGDG 551
QY 286 EAAADDALATMTDIDEATDGTAGVQLGERAADVESVTGVDDIDIAEQTNMLALNASTEA 345

Db 552 QAAQDAVAQLEIEDETEQAAATAVDDLEAKMSEIETIVAAITDIAEQTNMLALNASTEA 611
QY 346 ARAGEAGEGFAVVADEVKALAESEEQSTRVEELVEQMAETETVQDLQDEYNQRIEGCV 405
Db 612 ARADQDGDGFAVVADEVKDLADESKASAAETEAALVAEVAQQTETSVAAMDRIOERVSDGV 671
QY 406 ERVEEAMETLQBITDAVEDAASGMQEVSTATDEQAVSTEVEVAEMVDGVDDRAGEATAALD 465
Db 672 ETVSETERSLSIAGRIAEADTGVQEIINAMDDQAASVSDVTTAVGADVAAALGEETATEAE 731
QY 466 DIADATDQOQRTVEEV 481
Db 732 STADAAAEQASTLSDV 747

RESULT 7
HTR4_HALN1
ID_HTR4_HALN1 STANDARD; PRT; 810 AA.
AC Q9HP84;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Halobacterial transducer protein IV.
GN HTR5 OR HTPIV OR VNG1760G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA NG W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genomic sequence of Halobacterium species NRC-1.";
Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).
CC -!- FUNCTION: Potentially involved in chemo- or phototactic signal
transduction (By similarity)
CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC -!- SIMILARITY: Contains 2 HAMP domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DE EMBL; AE005080; AAC19986.1; -.
DR PIR; F84327; F84327.
DR HSSP; P02942; 1QU7.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMP; 2.
DR Transducer; Transmembrane; Complete proteome; Repeat.
KW DOMAIN 1 38 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 39 59 POTENTIAL.
FT DOMAIN 60 323 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 324 344 POTENTIAL.
FT DOMAIN 345 810 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 345 397 HAMP 1.

```


CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE005080; AAG19985.1; --
 DR EMBL; X95589; CAA64842.1; --
 DR PIR; E84327; E84327.
 DR PIR; T46811; T46811.
 DR HSSP; P02942; 1Q07.
 DR InterPro; IPR004089; Chmtaxis_trans.
 DR InterPro; IPR003660; HAMP.
 DR InterPro; IPR004090; Me.Chemotaxis.
 DR Pfam; PF00015; MCPsignal; 1.
 DR PRINTS; PR00260; CHEMTRNSDUCR.
 DR SMART; SM00304; HAMP; 1.
 DR SMART; SM00283; MA; 1.
 DR PROSITE; P55011; CHEMOTAXIS_TRANSDUC_2; 1.
 DR PROSITE; P55085; HAMP; 1.
 DR Transducer; Transmembrane; Complete proteome.
 FT TRANSMEM 10 30 POTENTIAL.
 FT TRANSMEM 44 64 POTENTIAL.
 FT TRANSMEM 80 100 POTENTIAL.
 FT DOMAIN 164 217 HAMP.
 FT DOMAIN 236 476 METHYL-ACCEPTING TRANSDUCER.
 SQ SEQUENCE 545 AA; 57070 MW; FDD870389C2F428B CRC64;

Query Match 21.8%; Score 521; DB 1; Length 545;

Best Local Similarity 29.6%; Pred. No. 1.3e-15;
 Matches 142; Conservative 94; Mismatches 195; Indels 48; Gaps 10;

QY 28 LDEAETAWLSFTGIDDDTMAALAAEQPLFEATADALVTDFF----YDHLESYERTQDLFA 93
 DB 64 LEKAGVATQI-FNLATGFLMTAVVFAIGGTATLAIEDMBEALVEERAQRQAEEREA 122
 QY 84 NSTKTYEQQLKETQAEYLLGLGRGEYDTEYAAQKRIKIHVDVLGLGPDVVLGAYTRYTG 143
 DB 123 EAERAREKAEQKQAE----AERQTAESAQKQDAR-ERSAEIQLAAD--LESQATEVGA 175
 QY 144 LLDALAD-----DVVARGE--EAAAADVELVARFLPMLKLLTPDQIAMDYIDSVA 194
 DB 176 TLEAASDGLTARVDATTONAEIAEVATVNDM-----LTTWERTI----- 216
 QY 195 QRLHDIIDSRQELANAVATHVEAPLSLEATSDQVAERTDTMRARDDQVDRMADYSRI 254
 DB 217 ----DSIQGFSTNTTASREATAKKEIQDASQTVSESVOEIAAGTDDQREQLSVAEEM 272
 QY 255 SVSASVEEVAETADVRRTSEDAEALAQGGAAADALATMTDIDEATDGTAGVEQLG 314
 DB 273 DSVSATVEEVAATAQSVADTAADTTDVATAGTQAEIDAIDAQVETWTTVANVDAL 332
 QY 315 ERAADVESVTGIDDIATQTNMLNANASIEARAGEA-----GEGFAVVADEKALAEER 370
 DB 333 DUTTRIDDIATLISDTAEQTNMLNANANIEARAGSGGSGSDGEGFAVVADEKALATESQ 392
 QY 371 EOSTVEELVEQMAETETVDQDDEVNORIGEGVERVEEAMETLQEIITDAVEDAASGMQ 430
 DB 393 RSAKDIAELIEVQSQTATTVEIRVAEQRVNDGAAAEVETVDAFGAVTNIQETDVGQ 452
 QY 431 EYSTATDEQAVSTEVEAEVMDGVDVDRAGIAAALDDIADATDQQRVTVEEVRTVGKLS 489
 DB 453 EISQAMDEQAQRSSRVSSVDDIATISQATADRAENVSAASEQTASITEVTSLSQLA 511

RESULT 11

ID HTR2 HALVA
 AC P42258;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis

DE protein II) (MPP-II) (Fragment).
 GN HTRII.
 OC Haloarcula vallismortis.
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Haloarcula.
 OX NCBI_TaxID=28442;
 RN [1]_SEQUENCE FROM N.A.
 RP STRAIN=ATCC 29715;
 RC MEDLINE=95224074; PubMed=7708770;
 RA Seidel R., Scharf B., Gautel M., Kleine K., Oesterhelt D.,
 RA Engelhard M.;
 RT "The primary structure of sensory rhodopsin II: a member of an
 RT additional retinal protein subgroup is coexpressed with its
 RT transducer, the halobacterial transducer of rhodopsin II.,"
 RL Proc. Natl. Acad. Sci. U.S.A. 92:3036-3040(1995).
 CC -!- FUNCTION: Transduces signals from the phototaxis receptor sensory
 CC rhodopsin II (SR-II) to the flagellar motor. Responds to light
 CC changes through the variation of the level of methylation. Also
 CC acts as a chemotransducer.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
 CC -!- SIMILARITY: Contains 1 HAMP domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; Z35308; CAA84549.1; --
 DR HSSP; P02942; 1Q07.
 DR InterPro; IPR004089; Chmtaxis_trans.
 DR InterPro; IPR003660; HAMP.
 DR InterPro; IPR004090; Me.Chemotaxis.
 DR Pfam; PF00015; MCPsignal; 1.
 DR PRINTS; PR00260; CHEMTRNSDUCR.
 DR SMART; SM00304; HAMP; 1.
 DR SMART; SM00283; MA; 1.
 DR PROSITE; P55011; CHEMOTAXIS_TRANSDUC_2; 1.
 DR PROSITE; P55085; HAMP; 1.
 DR Transducer; Photoreceptor; Chemotaxis; Transmembrane; Methylation.
 FT NON_TER 1 1
 FT DOMAIN 58 111 HAMP.
 FT DOMAIN 130 366 METHYL-ACCEPTING TRANSDUCER.
 SQ SEQUENCE 433 AA; 45935 MW; 90507B8897D943C0 CRC64;

Query Match 20.6%; Score 492; DB 1; Length 433;
 Best Local Similarity 29.9%; Pred. No. 1.8e-14;
 Matches 127; Conservative 86; Mismatches 172; Indels 40; Gaps 5;

QY 69 YDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQKRIKIHVDLGI 128
 DB 11 FDNMRANLRTQISEATAKQAEAAKEQAQ-----AAREDSERSEMEALTG----- 58
 QY 129 GPDVILGAYTRYTGLLDALADD-----VADRGEEAAAADVELVARFLPMLKLLTFDQ 183
 DB 59 ----HLEKAAQYSDALDAANGDLTARVKTDSMDMAEVEGDIINTTLDALEDTIADMK 114
 QY 184 IAMDVTIDSYAQRHDEIDSRQELANAVATHVEAPLSLEATSDQVAERTDTMRARDDQ 243
 DB 115 AFATNVIQS-----SDRVNSNAE-----RVDRASKQVSKSINEIFSGTTEQ 155
 QY 244 VDRMADVSRITSSVASVEEVASTADDVRRTSSEDAEALAQGGAAADALATMTDIDEAT 303
 DB 156 NEGLESAEAEMQNLGATAQQVASSAQVADTSSQAAKVGEDGREAQAQAEASATEART 215
 QY 304 DGVTAGVQLGERADVSTGVIDDIAEOTNMLNANASIEARAGEAGEGFAVVADEVK 363
 DB 216 GETVEEINALDDDELDEIGEIVGVITTSVEQTNMLNANASIEAHHADGDGEGFAVVADEIK 275

```

QY 364 ALAESREOSTRVEELVEQMAETEETVDQLDEVNORIGGEVERVEEAMETLQEIITDAVE 423
Db 276 GLAAETKEAADIIEGRIEALQAGQDTVETMESTSTRITEGTVSTVETVDALETIVYTE 335
QY 424 DRASGMQEVSTATDEQAVSTEEVAEMVGDVDRAGISAAALDDIADATDQQVTRVEEVR 483
Db 336 EVDTGIGIEDRATEEQARTAQDVMGTIDDLTTISQOTATEADTVAGAAQDSASIEVSD 395
QY 484 TVGKL 488
Db 396 SATEL 400

RESULT 12
HTR2 HALN1
ID HTR2 HALN1 STANDARD; PRT; 763 AA.
AC Q9HF61;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis
DE protein II) (MPP-II).
GN HTR2 OR VNGI765G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]_SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -!- FUNCTION: Transduces signals from the phototaxis receptor sensory
CC rhodopsin II (SR-II) to the flagellar motor. Responds to light
CC changes through the variation of the level of methylation. Also
CC acts as a chemotransducer (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC -!- SIMILARITY: Contains 2 HAMP domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005080; AAG19989.1; -.
DR PIR; A84328; A84328.
DR HSP; P02942; LQ07.
DR InterPro; IPR004089; Chntaxis_transd.
DR InterPro; IPR003650; HAMP.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS01111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS00885; HAMP; 2.
KW Transducer; Photoreceptor; Chemotaxis; Transmembrane; Methylation;
KW Complete proteome; Repeat.
FT INIT_MET 0 0 BY SIMILARITY.
FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 16 36 POTENTIAL.
FT DOMAIN 37 277 EXTRACELLULAR (POTENTIAL).

```

```

FT TRANSMEM 278 297 POTENTIAL.
FT DOMAIN 298 763 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 302 354 HAMP 1.
FT DOMAIN 396 449 HAMP 2.
FT DOMAIN 468 704 METHYL-ACCEPTING TRANSDUCER.
SQ SEQUENCE 763 AA; 78911 MW; CF7A8FF04DFF309A CRC64;

Query Match 19.4%; Score 464; DB 1; Length 763;
Best Local Similarity 28.9%; Pred. No. 4.8e-13;
Matches 136; Conservative 74; Mismatches 219; Indels 42; Gaps 9;

QY 37 LSFTGIDDDTMAALAAEQPLFEATDALVTDFVHLESYERTQDLFANSTKTKVEQLKETQ 96
Db 293 LSLIGTVGSTTVTALRQ--FSRRADEMAAGDLDTIDTSR-NDEGTTAEFSRWRDSL 349
QY 97 AEYLLGLGRGEYDTEYAAQARIGKIHVLGPDVYLGYTRYTYTGLLDALADVVADR 156
Db 350 SSSLT-----DAERATARAEDAR-EDAEQQRADA-----EAREDAEAAR 388
QY 157 --GEEAANAVDLVARFLPMLKLT---FDQIAMDITYDSYAQRLLH-----DEIDSQ 205
Db 389 KDAQETARALESAADYEALTAVDAGDLTRRVDASRDHDMARIGHALNDMLDDIETSV 448
QY 206 ELANAVATVHEAPLGLSLEA-----TSQDVAERTDTMRATDDQVDRMADYSREISSVS 258
Db 449 AATAFSDHVSDAQORVEADAGDAIDAGTDTVSTAVDEISDGATEQDRLHVEAGEVDDL 508
QY 259 ASVEEVASTADDVVRTSDEAALAOQGEAAADALATMTDIDEATDGVTVAGVEQLGERAA 318
Db 509 ASAEVETVSLADTAGAASAVDDGQATDEAVETMDVDADDAEAAAADAMDALDSEMA 568
QY 319 DVESVTGVDIDDAETNMALNASTAEARAGAGGFAVVADEVKALAESREOQSTRVEE 378
Db 569 DIGEIVDVADIADQTNMLNASTAEARTGADGDFVAVDEVKTLAESRDAEDIES 628
QY 379 LVEQMAETEETVDQLDEVNORIGGEVERVEEAMETLQEIITDAVEAAGMGEVSTATDB 438
Db 629 RLALQGGQVSDVADDEMRTSDTVSGRATVGDAATLDDVSVFVADTDTAAQEIRANTDR 688
QY 439 QAVSTEVAEMVGDVDRAGEIAAALDDIADATDQQVTRVEEVRVTKLS 489
Db 689 QAHAASRVASAVDEVAGISQETAQATAVADSAAATQDTLSSVDDAAAADLA 739

RESULT 13
HTR1 HALN1
ID HTR1 HALN1 STANDARD; PRT; 535 AA.
AC P33741; Q9HPP6;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sensory rhodopsin I transducer (HTR-I) (Methyl-accepting phototaxis
DE protein I) (MPP-I).
GN HTR1 OR HTR1 OR VNGI659G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081), and
OS Halobacterium halobium.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091; 2242;
RN [1]_SEQUENCE FROM N.A.
RX STRAIN=NRC-1;
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).

```


Search completed: August 10, 2004, 15:26:49
Job time : 16.3923 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:22:05 ; Search time 58.609 Seconds

(without alignments)
2632.505 Million cell updates/sec

Title: US-09-455-978B-2

Perfect score: 2394

Sequence: 1 MSNDNDTLVTADVRNGIDGH.....ATDQQVTVVEVRETVGKLS 489

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriaph:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Description
1	609.5	25.5	P71412
2	597	24.9	Q9HP10
3	597	24.9	Q9P9J0
4	580	24.2	Q9HR88
5	560.5	23.4	Q93775
6	540.5	22.6	P71415
7	539	22.5	P71414
8	535.5	22.4	Q9HQX8
9	529	22.1	P71416
10	524	21.9	O06022
11	523.5	21.9	P71409
12	523.5	21.9	Q9HPQ5
13	521.5	21.8	Q9HRN6
14	509.5	21.3	Q9HRA1
15	507.5	21.2	O93643
16	506	21.1	Q9HPW6

17	505	21.1	642	1	O59634
18	502.5	21.0	773	1	Q9UXS0
19	499.5	20.9	627	17	Q9HS86
20	497.5	20.8	419	1	O93644
21	497	20.8	536	17	Q9HPA2
22	475.5	19.9	537	16	Q9AB06
23	452.5	18.9	423	17	Q9HQT5
24	439.5	18.4	423	1	O93642
25	419.5	17.5	551	16	Q815U4
26	419.5	17.5	559	16	Q9K669
27	415.5	17.4	555	16	Q88L17
28	415	17.3	555	16	O9A5Y0
29	413	17.3	439	16	Q9KFH4
30	410.5	17.1	678	16	Q8P6P5
31	409.5	17.1	632	16	Q9HW93
32	408	17.0	499	16	Q8UGL0
33	405.5	16.9	579	16	Q9K632
34	405.5	16.9	632	2	O32440
35	405	16.9	637	16	Q8D3W6
36	403	16.8	539	16	Q87QG4
37	403	16.8	678	16	Q8PHZ9
38	401	16.8	530	16	Q9XOM7
39	401	16.8	629	16	Q9HW91
40	400.5	16.7	553	16	Q9KNF6
41	400	16.7	664	16	Q8RBV6
42	399.5	16.7	535	16	Q91OR3
43	399.5	16.7	563	16	Q81XF7
44	399	16.7	702	16	Q97D01
45	398	16.6	623	16	Q8D3R3

ALIGNMENTS

RESULT 1

P71412					
ID	P71412	PRELIMINARY;	PRT;	482	AA.
AC	P71412;				
DT	01-FEB-1997 (TREMBLrel. 02, Created)				
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)				
DT	01-JUN-2003 (TREMBLrel. 24, Last annotation update)				
DE	Transducer Hta protein.				
GN	HTA.				
OS	Halobacterium salinarum.				
OC	Archaea; Euryarchaeota; Halobacteria; Halobacteriales;				
OC	Halobacteriaceae; Halobacterium.				
OX	NCBI_TaxID=2242;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Flx15;				
RX	MEDLINE=96209786; PubMed=8643458;				
RA	Zhang W., Brooun A., McCandless J., Banda P., Alam M.;				
RT	"Signal transduction in the archaeon Halobacterium salinarum is processed through three subfamilies of 13 soluble and membrane-bound transducer proteins.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 93:4649-4654(1996).				
DR	EMBL; U75435; AAB17880.1; -.				
DR	PIR; T44973; T44973.				
DR	HSSP; P02942; 1QU7.				
DR	GO; GO:0016020; C:membrane; IEA.				
DR	GO; GO:0004871; F:signal transducer activity; IEA.				
DR	GO; GO:0005935; P:chemotaxis; IEA.				
DR	GO; GO:0007165; P:signal transduction; IEA.				
DR	InterPro; IPR004089; Chmtaxis_transd.				
DR	InterPro; IPR003660; HAMP.				
DR	InterPro; IPR004090; Me chemotaxis.				
DR	InterPro; IPR000014; PAS domain.				
DR	Pfam; PF00015; MCPsignal; 1.				
DR	PRINTS; PR00260; CHEMTRNSDUCR.				
DR	SMART; SM00304; HAMP; 1.				
DR	SMART; SM00283; NA; 1.				
DR	SMART; SM00091; PAS; 1.				
DR	PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.				

[illegible]

RESULT 5

```
O93775
ID O93775 PRELIMINARY; PRT; 452 AA.
AC O93775;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Car protein.
GN CAR.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=89;
RX MEDLINE=99164082; PubMed=10064592;
RA Storch K.F., Rudolph J., Oesterhelt D.;
RT "Car: A cytoplasmic sensor responsible for arginine chemotaxis in the
RT archaeon Halobacterium salinarum.";
RL EMBO J. 18:1146-1158(1999).
DR EMBL; AJ132321; CAB39318.1; -.
DR PIR; T44849; T44849.
DR HSSP; P02942; 1QU7.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis_trans.
DR InterPro; IPR004090; Me_chemotaxis.
DR InterPro; IPR00014; PAS domain.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00283; MA; 1.
DR TIGRFAMs; TIGR00229; sensory_box; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
SQ SEQUENCE 452 AA; 49098 MW; 9244D90D694681A CRC64;

Query Match 23.4%; Score 560.5; DB 1; Length 452;
Best Local Similarity 31.3%; Pred. No. 1.4e-16;
Matches 159; Conservative 88; Mismatches 170; Indels 91; Gaps 13;

QY 10 TADVRNGIDGHALADRI-----GLDEAEIA-----WRLGFTGIDDDTMAALAA 52
DB 5 SSDMGGEATGEHLADELCEAYLGDNEDDGGDELQRLSRERDFWKHFN-----QLVA 56

QY 53 EQP--LFEATADALVTD---FYDHLESY-----ERTQDLFANSTKTVEQLKETAQYL 100
DB 57 EYEGILITAAAGCTVTHWNERFSDHMKMARSDALGEDASDVFS---TABE-SETLPEAV 111

QY 101 LGLGRGEYDEYAAQARIGIKIHVGLGPDVYLGYTRYTYTGLLDALADDVVADRGEAA 160
DB 112 VRTGDTVEEEE-----PHDVPDLSLCQYHGVPLRAPTG-----DVVGSFG--V 152

QY 161 AAADVDELVARFLPMLKLLTFDQIQAMDYIDSVAQRLHDEIDSRQELANAVATHVEAPLS 220
DB 153 VPDISEKVKV-----QRELHD-----LHETVSSNVGEHLS 182

QY 221 SLEATSDVAERTDTTMRARTDDQVDRMADVSRISVSVAEVAETADVTRTSDEAAE 280
DB 183 ELSIESIDVGSFABETEAFAGEIERMEGVADSVNSQATIEBIASSAEVVSQASQAQD 242

QY 281 LAQCGEAAADDALATMTDIDEATDGTAGVEQLGERAAADVSTGVTDIDIAETQNMALN 340
DB 243 RATEGQTAFETADRMGAVQESAEVNDITDGLTSQADEMSEITIDAINDIADQTNMALN 302

QY 341 ASTEERARAGAGGFAVADVEKALAEESREQSTRVEELVEQMAETETVDQLDEVNQR 400
DB 303 ASTEERARAGKGGFAVADVEKSLAESQERADEIEQMVENVETTDQADRIGQTTIE 362

QY 401 IGGGVSRVEAMETLQBITDAVEDAAGMQEVSTATDEQAVSTEEVAEMVGVDDDRAGEI 460
DB 363 IEEAITAVRETLDSLQBIKRAVDETATGVKEVAGATDDHAASTEQVAAATDEAVDKLTEL 422
```

```
QY 461 AAALDDIADATDQOVRTVEVRETVGKL 488
DB 423 EDRLLNLSQIASSEQHQRVAEIEDMVDEL 450

RESULT 6
ID P71415 PRELIMINARY; PRT; 777 AA.
AC P71415;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Transducer Htd protein.
GN HTD.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=FX15;
RX MEDLINE=96209786; PubMed=8643458;
RA Zhang W., Brooun A., McCandless J., Banda P., Alam M.;
RT "Signal transduction in the archaeon Halobacterium salinarum is
RT processed through three subfamilies of 13 soluble and membrane-bound
RT transducer proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:4649-4654(1996).
DR EMBL; U75438; AAB17883.1; -.
DR PIR; T44597; T44597.
DR HSSP; P02942; 1QU7.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis_trans.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me_chemotaxis.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMP; 1.
SQ SEQUENCE 777 AA; 82097 MW; 0B20A6A593AA816 CRC64;

Query Match 22.6%; Score 540.5; DB 1; Length 777;
Best Local Similarity 32.4%; Pred. No. 1.8e-15;
Matches 141; Conservative 74; Mismatches 171; Indels 49; Gaps 6;

QY 88 TVEQLKETQAEVLLGLGRGEYDEYAAQRA-RIGKIHVGLGPDVYLGYTRYTYTGLLD 146
DB 320 TVRALNDLEAK-APALERGEYDTLDVARVDELGRLEFAFASLRDAVQVADSANEQQVD 378

QY 147 ALA-----DDVVADRGEEAAAADVDELVARFLP 173
DB 379 AEARSEAAQAQPEAAQAQAEAEAEESDAGAPGDDRRGVLGDDACVNAAGDLTVR--- 435

QY 174 MLKLLTFDQIIA----MDTYIDSVAQRLHDEIDSRQELANAVA---THVEAPLSLEATS 226
DB 436 ----LDADVEQAAMADMRRAFNEMAADMEATIAERGAFADEVATASTDASDSAAAVEQTG 491

QY 227 QDVAERTDTMARTDDQVDRMADVSRISVSVAEVAETADVTRTSDEAALAQOGE 286
DB 492 RDVSTAVGIRDRRAADQRLQLEAVASETDEMSATIEEVARVAGQVAETSORAALGDDGQ 551

QY 287 AAADDALATMTDIDEATDGTAGVEQLGERAAADVSTGVTDIDIAETQNMALNASTEAA 346
DB 552 AAAQDAVAQLEIEDETDQAAATAVDLLEAKMSEIETIVAAITDIAETQNMALNANIEAA 611

QY 347 RAGEAGEGFAVADVEKALAEESREQSTRVEELVEQMAETETVDQLDEVNQRIGEGVE 406
```

```

Db 612 RADQDGGFAVAVDEKDLADESKALAAEMQALVAEVRAQTETSVAAAMDRIQERVSDGVE 671
QY 407 RVEEMETLQEIITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVGVDDRAGRIAAALDD 466
Db 672 TVSETERSLDIAGRTAEPDGTQVEISNAMDDQRAASVSDVTTAVGVDVAALGEETATEAES 731
QY 467 IADATDQQVRTVEEV 481
Db 732 TRDAAAEQATTLSDV 746

RESULT 7
P71414
ID P71414 PRELIMINARY; PRT; 792 AA.
AC P71414;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Transducer Htc protein.
GN HTC.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Flx15;
RA Zhang W., Brooun A., McCandless J., Banda P., Alam M.;
RX MEDLINE=96209786; PubMed=8643458;
RT "Signal transduction in the archaeon Halobacterium salinarum is
RT processed through three subfamilies of 13 soluble and membrane-bound
RT transducer proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:4649-4654(1996).
DR EMBL; U75437; AAB17882.1; -.
DR HSSP; P02942; 1QU7.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me_chemotaxis.
DR Pfam; PF00672; HAMP; 1.
DR PRINTS; PR00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMP; 2.
SQ SEQUENCE 792 AA; 83664 MW; 745D3693F3EEF9F9 CRC64;

Query Match
Best Local Similarity 22.5%; Score 539; DB 1; Length 792;
Matches 143; Conservative 76; Mismatches 163; Indels 62; Gaps 7;

QY 42 IDDDTMAALAEQPLFEATADALVDFVDHLRSYERT----ODLEANSKTKVEQLKETQA 97
Db 374 LDELPLGAFGSLSQMH-RLEALITDLEARDAEQTRKDAEAAERLNERLERRAA 432
QY 98 EYLLGLGRGEYDTEVAQARIGKIHVLGLGPDVYLGYTRYTTGLLDALADDVVDRG 157
Db 433 -----EYSDENAAAA-----GDLTR-----RLDEDVDSGM 459
QY 158 EEAARAVDELVARFPLMKLLITFDQOIAMDITVDSYAQRLHDEISROELANAVATHVEA 217
Db 460 QDIAEAFNDMGDV-----EATLAQVRSTADAVDAATDVHDAABIR----- 501
QY 218 PLSSLEATSQDVAERTDTMTARTDDQVDRMADVRSISVSASVEVASTADDVPTSED 277
Db 502 -----SASDQVESVQDISADAEQRDLGLTGVGEVHALSATVEDIARPTSPSTVQ 555
QY 278 AEALAQCGEAAADDALATMTDIDEATDGTAGVEQLGERAADVESVTGVDDIDIAEQTNML 337

```

```

Db 556 AATESERGQELGEDAVAELERIETATDSAVERTALBEAVIDAIGHVTGVTITDIAQNML 615
QY 338 ALMASTAEAPAGAGGFAVAVDEKDLADESKALAAEMQALVAEVRAQTETSVAAAMDRIQERVSDGVE 397
Db 616 ALNANIEAARADKSGDGFVAVDEKDLADESKALAAEMQALVAEVRAQTETSVAAAMDRIQERVSDGVE 675
QY 398 NORIGEGVERVEEAMETLQEIITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVGVDDR 457
Db 676 GDRVDAGSETIEAALAAVDDIGDQVEAANGSVQISDATDEQAASTEEVVMTIDEVTDLS 735
QY 458 GEIAAALDDIADATDQQVRTVEEV 481
Db 736 DRTATESQQVSAAAEQQAASVSEV 759

RESULT 8
Q9HGX8
ID Q9HGX8 PRELIMINARY; PRT; 636 AA.
AC Q9HGX8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Htr15 transducer.
GN HTR15 OR VNG0958G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welter R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005032; AAG19381.1; -.
DR PIR; A84252; A84252.
DR HSSP; P02942; 1QU7.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR004090; Me_chemotaxis.
DR InterPro; IPR000700; PAS-associat.
DR Pfam; PF00015; MCPsignal; 1.
DR Pfam; PF00989; PAS; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00283; MA; 1.
DR SMART; SM00091; PAS; 1.
DR TIGRFAMs; TIGR00229; sensory_box; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50113; PAS; 1.
DR PROSITE; PS50112; PAS; 1.
KW Complete proteome.
SQ SEQUENCE 636 AA; 67348 MW; 98D2E415295D4D14 CRC64;

Query Match
Best Local Similarity 22.4%; Score 535.5; DB 17; Length 636;
Matches 149; Conservative 99; Mismatches 191; Indels 63; Gaps 11;

QY 17 IDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT----ADALVTD----- 67
Db 165 VDDALLDSIGMP-----VFMLDATG-----SVVANNETIELTGCCKADAMGMDHASEA 214
QY 68 FYDHLSEYRTQDLFANSTKTVQEQKETAQBYLLGLGRGEYDTEVAQARIGKIHVDL 127

```

```

Db 215 FYPDRVRKTLAD-----KVIESPRSA-----EFDIERSEKQALYADTSVM- 258
QY 128 LGPDVYLGYATRYTGLDALADD-----VVADGEEA--AAAVDELVARFLPML- 175
Db 259 ----TDGGTDRHIRFKASPIFDDDGELIAVAETIEDRTEDVRRADAVEELDELSTIID 314
QY 176 -----KLLTFDQIAMDYIDSYAQELHDEIDSRQELANAVATHVEAPLSSLEATSQ 227
Db 315 ALLSSQLSKRASFEHIGIINEQLVSVSALNGMADQFERLVGVQVGGQTOELADTIERATA 374
QY 228 DVAERTDTRKARTDQOVDMADVRSREISSVSASVEVASTADDVRRTSDBALAQOGEA 287
Db 375 DATDADTV-----SSQNEMLSSANEMENFASMQEVAASSQVSAAEQAQDAAESGLE 430
QY 288 AADDALATMTDIDEATDGTAGVQELGERAAVSVGTGVIDDIAEQTNNMALNASTAEAR 347
Db 431 ASEGANQATNEVIDISDDLMESVSKLESMDIEDVVEIAEVADQTNLLALNANIEAAR 490
QY 348 AGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMAETETVDCOLDDEVNORIGEVER 407
Db 491 AGEAGSGFAVVADEVKELANETREHTERFAGTSIDVQQQANETVLAVESHQIHRAGDE 550
QY 408 VEEAMETLQEIITDAVEDAASQMOEVSTATDEQAVSTEEVAEMVGVDDDRAGEITAAALDDI 467
Db 551 IDALTALAEIATSDVEAATGITEVARANDEQASTVEDVIVITIEDVQQQAEAEAAASDRI 610
QY 468 ADATDQOVRTVEEVRTVGKLS 489
Db 611 VSAEQEQSTAVSQLSRVDKLT 632

```

RESULT 9

```

P71416
ID P71416 PRELIMINARY; PRT; 804 AA.
AC P71416;
DT 01-FEB-1997 (T:EMBLrel. 02, Created)
DT 01-FEB-1997 (T:EMBLrel. 02, Last sequence update)
DT 01-JUN-2003 (T:EMBLrel. 24, Last annotation update)
DE Transducer Htf protein.
GN HTF.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FLX15;
RX MEDLINE=96209786; PubMed=8643458;
RA Zhang W., Brooun A., McCandless J., Banda P., Alam M.;
RT "Signal transduction in the archaeon Halobacterium salinarum is
RT processed through three subfamilies of 13 soluble and membrane-bound
RT transducer proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:4649-4654 (1996).
DR EMBL; U75439; AAB17884.1; -.
DR PIR; T44606; T44606.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis_trans.
DR InterPro; IPR003660; HAMP.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS01111; CHEMOTAXIS_TRANS2; 1.
DR PROSITE; PS00885; HAMP; 1.
SQ SEQUENCE 804 AA; 84705 MW; 0D57284683BFD18B CRC64;

```

Query Match 22.1%; Score 529; DB 1; Length 804;
 Best Local Similarity 30.6%; Pred. No. 5.9e-15;
 Matches 139; Conservative 76; Mismatches 159; Indels 80; Gaps 9;

```

QY 89 VEOLKETQAEYLLGLGRGEYDTEVAAQR-----ARIGKIHVGLGPDVYLGYATRY 140
Db 351 VQSLSAARAE-----LEAGNYDQVATSRDEIGQLFASIGSMRDALVTQAE--AAREQA 403
QY 141 YTGLLDALAD-----DVADRGEEAAAADVELVARFLPMLKLLTFDQOIAMDYIDSYAQR 196
Db 404 TEAQDAEADARRAEDARAEDAKADAEALAAE-----LEQAQR 444
QY 197 LHDEI-----DSR-----QELANAVAT---HV 215
Db 445 YSDVTGGVADGDLTRMPADDTNEMAAIAASFNDQWEHTIIDIQEPADAVATASEEA 504
QY 216 EAPLSLSLATSQDVAERTDTRMEARTDDQVDRMADVRSREISSVSASVEEVASTADDVRRTS 275
Db 505 EVGAADAERASQVSESQVEIAGAADQERNMLDVTSGEMTDLSSAAIEEVAAS-ADSAEHS 563
QY 276 EDAEALAQGEAAADALATMTDIDEATDGTAGVQELGERAAVSVGTGVIDDIAEQTN 335
Db 564 HQTAETARDGEQTAEDAIERSLTVQEAIDATVQNVLEALDDQMAEISEIVDLISDIAEQTN 623
QY 336 MLALNASTAEARAGEGFAVVADEVKALAEESREQSTRVEELVEQMAETETVDCOLD 395
Db 624 MLALNANIEAARADKSGGFAVVADEVKDFABEETQESAGDIERRITEVOSQTATVAEAR 683
QY 396 EYNORIGEVERVEEAMETLQEIITDAVEDAASQMOEVSTATDEQAVSTEEVAEMVGVDD 455
Db 684 AAESMDAGIDAVEEVVDAFTAVSDHSDETDGQVEISDTDDQAASTEAEVSMTEEVAD 743
QY 456 RAGEITAAALDDIADATDQOVRTVEEVRTVGKLS 489
Db 744 LSDSTAGEAQSVSATAEQAASMSSEISDTVESLS 777

```

RESULT 10

```

O06022
ID O06022 PRELIMINARY; PRT; 451 AA.
AC O06022;
DT 01-JUL-1997 (T:EMBLrel. 04, Created)
DT 01-JUL-1997 (T:EMBLrel. 04, Last sequence update)
DT 01-JUN-2003 (T:EMBLrel. 24, Last annotation update)
DE Soluble transducer protein Hth.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FLX15;
RX MEDLINE=97284501; PubMed=9139915;
RA Brooun A., Zhang W., Alam M.;
RT "Primary structure and functional analysis of the soluble transducer
RT protein HtrXI in the Archaeon Halobacterium salinarum.";
RL J. Bacteriol. 179:2963-2968 (1997).
DR EMBL; U74668; AAC45264.1; -.
DR PIR; T44964; T44964.
DR HSP; P02942; 1QU7.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis_trans.
DR InterPro; IPR004090; Me_chemotaxis.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF00015; MCPsignal_1.
DR PRINTS; PR00260; CHEMTRNSDUCH.
DR SMART; SM00283; MA; 1.
DR SMART; SM00091; PAS; 1.
DR TIGRfams; TIGR00229; sensory_box; 1.
DR PROSITE; PS01111; CHEMOTAXIS_TRANS2; 1.
SQ SEQUENCE 451 AA; 49122 MW; 6DA9E4F8139DA87 CRC64;

```

Query Match

21.9%; Score 524; DB 1; Length 451;

[illegible]

Qy 439 QAVSTEEVAEMVCGYDDRAGEIAAALDDIADATDQOVRTVEEV 481
Db 702 QADATQSVVRRYDDVADISQHVTEDAEQVSAAAEEQASVAEI 744

Search completed: August 10, 2004, 15:28:39
Job time : 60.609 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:18:30 ; Search time 26.4879 Seconds
(without alignments)
1632.058 Million cell updates/sec

Title: US-09-455-978B-76

Perfect score: 794

Sequence: 1 VLSEGEWQLVHWAKVEAD.....ALELPRKDIAAKYKELGYQG 153

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	794	100.0	153	2 AAY29391	Aay29391 Sperm wha
2	794	100.0	153	2 AAW81769	Aaw81769 Whale MYG
3	794	100.0	153	3 AAY69975	Aay69975 MYGL prot
4	715	90.1	154	2 AAW62270	Aaw62270 Modified
5	713	89.8	154	2 AAW29741	Aaw29741 Modified
6	711	89.5	154	2 AAW29739	Aaw29739 Modified
7	711	89.5	154	2 AAW29740	Aaw29740 Modified
8	711	89.5	154	2 AAW62271	Aaw62271 Modified
9	707	89.0	154	2 AAW29742	Aaw29742 Modified
10	701	88.3	154	2 AAW29743	Aaw29743 Modified
11	700	88.2	154	2 AAW62272	Aaw62272 Modified
12	699	88.0	154	2 AAW29744	Aaw29744 Modified
13	699	88.0	166	7 ADC31106	Aad31106 Human nov
14	699	88.0	206	7 ADC32817	Aad32817 Human nov
15	697	87.8	154	2 AAW62273	Aaw62273 Modified
16	613	77.2	202	4 ABG04708	Abg04708 Novel hum
17	535	67.4	231	4 ABG04706	Abg04706 Novel hum
18	472	59.4	113	3 AAG00022	Aag00022 Human sec
19	386.5	48.7	334	4 ABG21418	Abg21418 Novel hum
20	314	39.5	71	3 AAG00021	Aag00021 Human sec
21	253	31.9	107	4 ABG21417	Abg21417 Novel hum
22	243	30.6	64	2 AAR95437	Aar95437 Apomyoglo
23	240	30.2	60	2 AAR95436	Aar95436 Apomyoglo
24	215.5	27.1	190	5 ABB07967	Abb07967 Rat STUF
25	215.5	27.1	201	2 AAW86334	Aaw86334 Kidney in

ALIGNMENTS

RESULT 1

AAY29391
ID AAY29391 standard; protein; 153 AA.

XX AC AAY29391;
XX DT 01-OCT-1999 (first entry)
XX DE Sperm whale myoglobin protein sequence.
XX KW Myoglobin; Mb; sperm whale; protein design; function; property;
XX KM stereo structure; globular protein; detection.
XX OS Physeter sp.
XX PN JP11193297-A.
XX PD 21-JUL-1999.
XX PF 06-OCT-1998; 98JP-00283852.
XX PR 06-OCT-1997; 97JP-00272431.
XX (RIKA) RIKAGAKU KENKYUSHO.
XX WPI; 1999-462430/39.
XX Method for designing artificial protein - useful for producing proteins with required functions.
XX Example; Page 4-5; 10pp; Japanese.

The present invention describes a method for producing an ideal amino acid sequence. The method comprises: (1) preparation of one initial amino acid sequence corresponding to the structure of a protein; (2) selecting the amino acid residues optimal to each site; (3) repeating the steps of selecting the amino acid residues optimum to each sites of the total amino acid residues constituting the N-order amino acid sequence (N is an integer not less than 2) to give an N+1-order amino acid sequence consisting of the selected amino acid residues until the N-order amino acid sequence comes to be same as the N+1-order amino acid sequence; and (4) selecting the resultant N-order amino acid sequence as the optimum amino acid sequence of said protein. The method can design a protein with desired functions and properties. The present sequence represents sperm whale myoglobin used in an example from the present invention where the stereo structure of sperm whale myoglobin was targeted as a globular protein to detect an amino acid sequence best fit to the structure of the main chain of the sperm whale myoglobin

26	213	26.8	41	4	AAE04661	Aae04661 Haem prot
27	213	26.8	57	2	AAR95435	Aar95435 Apomyoglo
28	212.5	26.8	190	5	ABB07966	Abb07966 Human STU
29	212.5	26.8	190	7	ADC10192	Adc10192 Human NOV
30	212.5	26.8	190	7	ADC10194	Adc10194 Human NOV
31	204	25.7	41	4	AAE04662	Aae04662 Haem prot
32	203	25.6	41	4	AAE04665	Aae04665 Haem prot
33	202	25.4	41	4	AAE04664	Aae04664 Haem prot
34	198	24.9	39	4	AAE04669	Aae04669 Haem prot
35	198	24.9	39	4	AAE04670	Aae04670 Haem prot
36	194	24.4	39	4	AAE04673	Aae04673 Haem prot
37	191	24.1	39	4	AAE04672	Aae04672 Haem prot
38	191	24.1	41	4	AAE04663	Aae04663 Haem prot
39	189.5	23.9	153	2	AAY29393	Aay29393 Sperm wha
40	178	22.4	39	4	AAE04671	Aae04671 Haem prot
41	171.5	21.6	188	3	AAB41929	Aab41929 Human ORF
42	170	21.4	39	4	AAE04674	Aae04674 Haem prot
43	164	20.7	51	4	ABG04707	Abg04707 Novel hum
44	164	20.7	51	4	ABG04705	Abg04705 Novel hum
45	163	20.5	141	2	AAW73385	Aaw73385 Human hae

```

XX SQ Sequence 153 AA;
Query Match 100.0%; Score 794; DB 2; Length 153;
Best Local Similarity 100.0%; Pred. No. 3e-80;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKMSKASED 60
Db 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKMSKASED 60

QY 61 LKXHGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISBAIIHVLHSRHP 120
Db 61 LKXHGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISBAIIHVLHSRHP 120

QY 121 GDFGADAQAGAMNKALELFRKDIKAAKYKELGYQG 153
Db 121 GDFGADAQAGAMNKALELFRKDIKAAKYKELGYQG 153

RESULT 2
AAW81769
ID AAW81769 standard; peptide; 153 AA.
XX
AC AAW81769;
XX
XX 29-JAN-1999 (first entry)
XX
DE Whale MYGL peptide fragment.
XX
KW DHFR; dihydrofolic acid reductase; protein function; trypsin; bovine;
KW ribonuclease; myoglobin; database; homology; resemblance; whale.
XX
OS Cetacea.
XX
XX JP10287696-A.
XX
PD 27-OCT-1998.
XX
XX 11-APR-1997; 97JP-00093577.
XX
XX 11-APR-1997; 97JP-00093577.
XX
PA (IYAK-) IYAKU BUNSHI SEKKEI KENKYUSHO KK.
XX
XX WPI; 1999-018384/02.
XX
XX Determination of protein biological function - comprises use of amino
XX acid sequences database containing the relevant information.
XX
XX Example 1; Fig 1; 11pp; Japanese.
XX
XX This sequence is used in the creation of a database containing the
XX information for amino acid sequence of protein with at least 1 biological
XX function with added a score on importance of expression of the biological
XX information for each amino acid residue. The database is useful for
XX determination of unknown biological function of a protein or polypeptide
XX based on the homology of amino acid sequence, e.g. steric structure of
XX protein, and includes retrieval and evaluation of high homologous
XX relationship for the determination of mostly resembling protein. The
XX database allows for correct and rapid retrieval and presumption of
XX protein and polypeptide having biological functions

XX SQ Sequence 153 AA;
Query Match 100.0%; Score 794; DB 2; Length 153;
Best Local Similarity 100.0%; Pred. No. 3e-80;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKMSKASED 60
Db 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKMSKASED 60

```

```

QY 61 LKXHGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISBAIIHVLHSRHP 120
Db 61 LKXHGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISBAIIHVLHSRHP 120

QY 121 GDFGADAQAGAMNKALELFRKDIKAAKYKELGYQG 153
Db 121 GDFGADAQAGAMNKALELFRKDIKAAKYKELGYQG 153

RESULT 3
AAY69975
ID AAY69975 standard; protein; 153 AA.
XX
AC AAY69975;
XX
XX 13-APR-2000 (first entry)
XX
DE MYGL protein.
XX
KW Protein sequence database; biological function determination;
KW enzymatic activity; signaling activity; protein function determination;
KW MYGL protein.
XX
XX Unidentified.
XX
XX WO9962004-A1.
XX
XX 02-DEC-1999.
XX
XX 26-MAY-1998; 98WO-JP002302.
XX
XX 26-MAY-1998; 98WO-JP002302.
XX
XX (MEDI-) INST MEDICINAL MOLECULAR DESIGN INC.
XX
XX Itai A, Tomioka N, Itai R, Imamura M;
XX WPI; 2000-136797/12.
XX
XX An efficient, accurate and rapid computer database for estimating protein
XX functions e.g. enzymatic activity, for polypeptides obtained from gene
XX sequence translation.
XX
XX Example 1; Fig 1; 26pp; Japanese.
XX
XX This sequence represents the MYGL protein. The invention relates to a
XX database containing information on the amino acid (aa) sequences of
XX proteins of which 1 or more biological functions are known. The database
XX also contains additional information on the score of importance of each
XX aa residue in the whole aa sequence in determining the known biological
XX functions. The invention also relates to a method of preparing an
XX alignment between aa sequences contained in the database and those of the
XX unknown polypeptide. This is represented as the homology amongst various
XX sites, each being identified as having a high score of importance in
XX determining potential biological functions. The method is used to enable
XX an efficient estimation of the biological functions (particularly
XX enzymatic and signaling activities) of polypeptides from their aa
XX sequences. Suitable proteins can then be isolated and purified by various
XX means. This could be of considerable use in a biological and medical
XX context. The computerised procedure is efficient, fast and accurate
XX
XX SQ Sequence 153 AA;
Query Match 100.0%; Score 794; DB 3; Length 153;
Best Local Similarity 100.0%; Pred. No. 3e-80;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKMSKASED 60
Db 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKMSKASED 60

QY 61 LKXHGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISBAIIHVLHSRHP 120

```


Db	3	LSDGEWQVNLVNWGKVEADIAGHQEVLRIRLFTGHPETLEKFDKFKHLKTEAEKASDDL	62
Qy	62	KKHGVTVTLTALGAILKKKGHEAEKPLAQSHATKHKIPIKYLEFTISEAIHVLHSRHPG	121
Db	63	KKHGVTVTLTALGAILKKKGHEAEKPLAQSHATKHKIPIKYLEFTISEAIHVLHSKHPG	122
Qy	122	DFGADAQGMNKALELFRKDIKAAKYKELGYQG	153
Db	123	DFGADAQGMNKALELFRNDIAAKYKELGFQG	154
RESULT 7			
AAW29740	standard; protein; 154 AA.		
ID	AAW29740		
XX	AAW29740;		
AC	AAW29740;		
XX	26-OCT-1998 (first entry)		
DT	26-OCT-1998 (first entry)		
XX	Modified myoglobin protein 6.		
DE	Modified myoglobin protein 6.		
XX	Heart; myoglobin; Mb; random mutagenesis; peroxidase activity; Fe(III);		
KW	Mn(II) peroxidase activity; hydrogen peroxide; Mn binding site; catalyze;		
KW	phenolic; polymer; antigen; antibody; diagnostic assay.		
OS	Equus sp.		
OS	Synthetic.		
XX	CA2185605-A.		
PN	CA2185605-A.		
XX	17-MAR-1998.		
PD	17-MAR-1998.		
XX	16-SEP-1996; 96CA-02185605.		
PF	16-SEP-1996; 96CA-02185605.		
XX	16-SEP-1996; 96CA-02185605.		
PR	16-SEP-1996; 96CA-02185605.		
XX	(MAUK/) MAUK A G.		
PA	(WANL/) WAN L.		
PA	(LEEH/) LEE H.		
PA	(BRAY/) BRAYER G D.		
PA	(TONG/) TONG H.		
PA	(SMIT/) SMITH M.		
XX	Wan L, Mauk AG, Lee H, Brayer GD, Tong H, Smith M;		
PI	Wan L, Mauk AG, Lee H, Brayer GD, Tong H, Smith M;		
XX	WPI; 1998-388655/34.		
DR	WPI; 1998-388655/34.		
XX	New modified myoglobin with amino acid alterations has increased		
PT	peroxidase activity - used for catalysing oxidation of substrates with		
PT	peroxides, e.g. in waste treatment and as label for antigens.		
XX	Claim 10; Page; 23pp; English.		
PS	Claim 10; Page; 23pp; English.		
XX	Sequences AAW62270-W62273, and AAW29739-W29744, are modified synthetic		
CC	horse heart myoglobin (Mb) proteins, produced by random mutagenesis of		
CC	the synthetic Mb gene. This present sequence contains a substitution at		
CC	Thr391le which has been shown to enhance the peroxidase activity of Mb.		
CC	These protein can be used to catalyze oxidation of a substrate by		
CC	peroxidase, e.g. oxidation of Mn2+. Peroxidases are used for oxidising		
CC	phenolics to polymers, as labels for e.g. antigens, antibodies or		
CC	oligonucleotides, in diagnostic assays and for oxidative treatment of		
CC	waste streams. NB. This sequence is not given in the specification but		
CC	was created from the wild type sequence given by the inventors. The		
CC	position of the substitution given does not include the initiation codon		
XX	Sequence 154 AA;		
SQ	Sequence 154 AA;		
Query Match			
Best Local Similarity 89.5%; Score 711; DB 2; Length 154;			
Matches 133; Conservative 9; Mismatches 10; Indels 0; Gaps 0;			
Qy	2	LSEGEWQLVHWAKVADVAGHQDILIRLFKSHPETLEKFDKFKHLKTEAEKASDDL	61

Db 3 LSGEQVQLVNWGVKVEADVAGHGQDILIRLFKSHPETLEKFKDLKHLKTEAEMKASEDL 62
QY 62 KHGVTVLTVLALGAILKKKGHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSHRHPG 121
Db 63 KHGVTVLTVLALGAILKKKGHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSHRHPG 122
QY 122 DFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
Db 123 DFGADAQAGAMTKALELFRNDIAAKYKELGFQG 154
RESULT 8
AAW62271
ID AAW62271 standard; protein; 154 AA.
AC AAW62271;
XX
DT 26-OCT-1998 (first entry)
DE Modified myoglobin protein 2.
DE
KW Heart; myoglobin; Mb; random mutagenesis; peroxidase activity; Fe(III);
KW Mn(II) peroxidase activity; hydrogen peroxide; Mn binding site; catalyse;
KW phenolic; polymer; antigen; antibody; diagnostic assay.
XX
OS Equus sp.
OS Synthetic.
XX
PN CA2185605-A.
XX
PD 17-MAR-1998.
XX
PF 16-SEP-1996; 96CA-02185605.
XX
PR 16-SEP-1996; 96CA-02185605.
XX
PA (MAUK/) MAUK A G.
PA (WANL/) WAN L.
PA (LEEH/) LEE H.
PA (BRAY/) BRAYER G D.
PA (TONG/) TONG H.
PA (SMIT/) SMITH M.
XX
PI Wan L, Mauk AG, Lee H, Brayer GD, Tong H, Smith M;
XX WPI; 1998-388655/34.
XX
PT New modified myoglobin with amino acid alterations has increased
PT peroxidase activity - used for catalysing oxidation of substrates with
PT peroxides, e.g. in waste treatment and as label for antigens.
XX
PS Claim 6; Page; 23pp; English.
XX
CC Sequences AAW62270-W62273, and AAW29739-W29744, are modified synthetic
CC horse heart myoglobin (Mb) proteins, produced by random mutagenesis of
CC the synthetic Mb gene. This present sequence contains a substitution at
CC Lys45Glu, which has been found to enhance the peroxidase activity of the
CC protein in general, and it may also exhibit enhanced Mn(II) peroxidase
CC activity as well. This is due to the enhanced rate of reaction of Fe(III)
CC Mb with hydrogen peroxide and also because the substitution at residue 45
CC creates a new Mn binding site which is stronger than the natural Mn
CC binding site of Mb. This protein also contains a substitution at
CC Lys63Glu, which was also found to be required for Mn binding. These
CC protein can be used to catalyse oxidation of a substrate by peroxidase,
CC e.g. oxidation of Mn2+. Peroxidases are used for oxidising phenolics to
CC polymers, as labels for e.g. antigens, antibodies or oligonucleotides, in
CC diagnostic assays and for oxidative treatment of waste streams. NB. This
CC sequence is not given in the specification but was created from the wild
CC type sequence given by the inventors. The positions of the substitutions
CC given do not include the initiation codon
XX Sequence 154 AA;

Query Match 89.5%; Score 711; DB 2; Length 154;
Best Local Similarity 87.5%; Pred. No. 5.5e-71;
Matches 133; Conservative 9; Mismatches 10; Indels 0; Gaps 0;
QY 2 LSGEQVQLVHVMKVEADVAGHGQDILIRLFKSHPETLEKFKDLKHLKTEAEMKASEDL 61
Db 3 LSGEQVQLVNWGVKVEADVAGHGQDILIRLFKSHPETLEKFKDLKHLKTEAEMKASEDL 62
QY 62 KHGVTVLTVLALGAILKKKGHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSHRHPG 121
Db 63 KHGVTVLTVLALGAILKKKGHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSHRHPG 122
QY 122 DFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
Db 123 DFGADAQAGAMTKALELFRNDIAAKYKELGFQG 154
RESULT 9
AAW29742
ID AAW29742 standard; protein; 154 AA.
XX
AC AAW29742;
XX
DT 26-OCT-1998 (first entry)
DE Modified myoglobin protein 8.
DE
KW Heart; myoglobin; Mb; random mutagenesis; peroxidase activity; Fe(III);
KW Mn(II) peroxidase activity; hydrogen peroxide; Mn binding site; catalyse;
KW phenolic; polymer; antigen; antibody; diagnostic assay.
XX
OS Equus sp.
OS Synthetic.
XX
PN CA2185605-A.
XX
PD 17-MAR-1998.
XX
PF 16-SEP-1996; 96CA-02185605.
XX
PR 16-SEP-1996; 96CA-02185605.
XX
PA (MAUK/) MAUK A G.
PA (WANL/) WAN L.
PA (LEEH/) LEE H.
PA (BRAY/) BRAYER G D.
PA (TONG/) TONG H.
PA (SMIT/) SMITH M.
XX
PI Wan L, Mauk AG, Lee H, Brayer GD, Tong H, Smith M;
XX WPI; 1998-388655/34.
XX
PT New modified myoglobin with amino acid alterations has increased
PT peroxidase activity - used for catalysing oxidation of substrates with
PT peroxides, e.g. in waste treatment and as label for antigens.
XX
PS Claim 10; Page; 23pp; English.
XX
CC Sequences AAW62270-W62273, and AAW29739-W29744, are modified synthetic
CC horse heart myoglobin (Mb) proteins, produced by random mutagenesis of
CC the synthetic Mb gene. This present sequence contains substitutions at
CC Phe46Leu and Ile107Phe which have been shown to enhance the peroxidase
CC activity of Mb. These protein can be used to catalyse oxidation of a
CC substrate by peroxidase, e.g. oxidation of Mn2+. Peroxidases are used for
CC oxidising phenolics to polymers, as labels for e.g. antigens, antibodies
CC or oligonucleotides, in diagnostic assays and for oxidative treatment of
CC waste streams. NB. This sequence is not given in the specification but
CC was created from the wild type sequence given by the inventors. The
CC positions of the substitutions given do not include the initiation codon
XX Sequence 154 AA;

Query Match 89.0%; Score 707; DB 2; Length 154;
 Best Local Similarity 86.8%; Pred. No. 1.5e-70;
 Matches 132; Conservative 9; Mismatches 11; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASEDL 61
 DB 3 LSDGEWQVNLVWGVKVEADIAHGQEVLRIRLFTGHPETLEKFDKFKHLKTEAEKASEDL 62

QY 62 KKHGTVTLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIHVHLSRHPG 121
 DB 63 KKHGTVTLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIHVHLSRHPG 122

QY 122 DFGADAQAGAMNKALELFRKIDIAAKYKELGYQG 153
 DB 123 DFGADAQAGAMTKALELFRNDIAAKYKELGYQG 154

RESULT 10
 AAW29743
 ID AAW29743 standard; protein; 154 AA.
 XX
 AC AAW29743;
 XX
 DT 26-OCT-1998 (first entry)
 XX
 DE Modified myoglobin protein 9.
 XX
 KW Heart; myoglobin; Mb; random mutagenesis; peroxidase activity; Fe(III);
 XX Mn(II) peroxidase activity; hydrogen peroxide; Mn binding site; catalyse;
 KW phenolic; polymer; antigen; antibody; diagnostic assay.
 XX
 OS Equus sp.
 OS Synthetic.
 XX
 PN CA2185605-A.
 XX
 PD 17-MAR-1998.
 XX
 PF 16-SEP-1996; 96CA-02185605.
 XX
 PR 16-SEP-1996; 96CA-02185605.
 XX
 PA (MAUK/) MAUK A G.
 PA (WANL/) WAN L.
 PA (LEEH/) LEE H.
 PA (BRAY/) BRAYER G D.
 PA (TONG/) TONG H.
 PA (SMIT/) SMITH M.
 XX
 PI Wan L, Mauk AG, Lee H, Brayer GD, Tong H, Smith M;
 XX
 DR WPI; 1998-388655/34.
 XX
 PT New modified myoglobin with amino acid alterations has increased
 PT peroxidase activity - used for catalysing oxidation of substrates with
 PT peroxides, e.g. in waste treatment and as label for antigens.
 XX
 PS Claim 12; Page; 23pp; English.
 XX
 CC Sequences AAW62270-W62273, and AAW29739-W29744, are modified synthetic
 CC horse heart myoglobin (Mb) proteins, produced by random mutagenesis of
 CC the synthetic Mb gene. This present sequence contains substitutions at
 CC Thr35Ile, Phe46Leu, and Ile107Phe, which have been shown to enhance the
 CC peroxidase activity of Mb. These protein can be used to catalyse
 CC oxidation of a substrate by peroxidase, e.g. oxidation of Mn2+.
 CC Peroxidases are used for oxidising phenolics to polymers, as labels for
 CC e.g. antigens, antibodies or oligonucleotides, in diagnostic assays and
 CC for oxidative treatment of waste streams. NB. This sequence is not given
 CC in the specification but was created from the wild type sequence given by
 CC the inventors. The positions of the substitutions given do not include
 CC the initiation codon
 XX Sequence 154 AA;
 SQ

Query Match 88.3%; Score 701; DB 2; Length 154;
 Best Local Similarity 86.2%; Pred. No. 7.2e-70;
 Matches 131; Conservative 9; Mismatches 12; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASEDL 61
 DB 3 LSDGEWQVNLVWGVKVEADIAHGQEVLRIRLFTGHPETLEKFDKFKHLKTEAEKASEDL 62

QY 62 KKHGTVTLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIHVHLSRHPG 121
 DB 63 KKHGTVTLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIHVHLSRHPG 122

QY 122 DFGADAQAGAMNKALELFRKIDIAAKYKELGYQG 153
 DB 123 DFGADAQAGAMTKALELFRNDIAAKYKELGYQG 154

RESULT 11
 AAW62272
 ID AAW62272 standard; protein; 154 AA.
 XX
 AC AAW62272;
 XX
 DT 26-OCT-1998 (first entry)
 XX
 DE Modified myoglobin protein 3.
 XX
 KW Heart; myoglobin; Mb; random mutagenesis; peroxidase activity; Fe(III);
 XX Mn(II) peroxidase activity; hydrogen peroxide; Mn binding site; catalyse;
 KW phenolic; polymer; antigen; antibody; diagnostic assay.
 XX
 OS Equus sp.
 OS Synthetic.
 XX
 PN CA2185605-A.
 XX
 PD 17-MAR-1998.
 XX
 PF 16-SEP-1996; 96CA-02185605.
 XX
 PR 16-SEP-1996; 96CA-02185605.
 XX
 PA (MAUK/) MAUK A G.
 PA (WANL/) WAN L.
 PA (LEEH/) LEE H.
 PA (BRAY/) BRAYER G D.
 PA (TONG/) TONG H.
 PA (SMIT/) SMITH M.
 XX
 PI Wan L, Mauk AG, Lee H, Brayer GD, Tong H, Smith M;
 XX
 DR WPI; 1998-388655/34.
 XX
 PT New modified myoglobin with amino acid alterations has increased
 PT peroxidase activity - used for catalysing oxidation of substrates with
 PT peroxides, e.g. in waste treatment and as label for antigens.
 XX
 PS Claim 23; Page; 23pp; English.
 XX
 CC Sequences AAW62270-W62273, and AAW29739-W29744, are modified synthetic
 CC horse heart myoglobin (Mb) proteins, produced by random mutagenesis of
 CC the synthetic Mb gene. This present sequence contains a substitution at
 CC Lys45Glu, which has been found to enhance the peroxidase activity of the
 CC protein in general, and it may also exhibit enhanced Mn(II) peroxidase
 CC activity as well. This is due to the enhanced rate of reaction of Fe(III)
 CC Mb with hydrogen peroxide and also because the substitution at residue 45
 CC creates a new Mn binding site which is stronger than the natural Mn
 CC binding site of Mb. This protein also contains a substitution at
 CC Lys63Glu, which was also found to be required for Mn binding, and a
 CC substitution at His97Leu to increase the rate of Mn(II) turnover. These
 CC protein can be used to catalyse oxidation of a substrate by peroxidase,
 CC e.g. oxidation of Mn2+. Peroxidases are used for oxidising phenolics to

CC polymers, as labels for e.g. antigens, antibodies or oligonucleotides, in
CC diagnostic assays and for oxidative treatment of waste streams. NB. This
CC sequence is not given in the specification but was created from the wild
CC type sequence given by the inventors. The positions of the substitutions
CC given do not include the initiation codon
XX
SQ Sequence 154 AA;

Query Match 88.2%; Score 700; DB 2; Length 154;
Best Local Similarity 86.8%; Pred. No. 9.3e-70;
Matches 132; Conservative 9; Mismatches 11; Indels 0; Gaps 0;

QY 2 LSEGEQVLVHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASEDL 61
Db 3 LSDGEWQVNLVWVGKVEADIAHGQEVLRFTGHPELTLEKFDKFKHLKTEAEKASEDL 62

QY 62 KKHGTVLVTALGAILKKKGHEALKPLAQSHATKHKIPKYLEFTSEALIIHVLHSHRHPG 121
Db 63 KKHGTVLVTALGAILKKKGHEALKPLAQSHATKHKIPKYLEFTSEALIIHVLHSHRHPG 122

QY 122 DFGADAQAGAMTKALELFRNDIAAKYKELGYQG 153
Db 123 DFGADAQAGAMTKALELFRNDIAAKYKELGYQG 154

RESULT 12
AAW29744
ID AAW29744 standard; protein; 154 AA.

AC AAW29744;

XX 26-OCT-1998 (first entry)

XX Modified myoglobin protein 10.

XX Heart; myoglobin; Mb; random mutagenesis; peroxidase activity; Fe(III);
KW Mn(II) peroxidase activity; hydrogen peroxide; Mn binding site; catalyse;
KW phenolic; polymer; antigen; antibody; diagnostic assay.

XX Equus sp.

OS Synthetic.

XX CA2185605-A.

PN 17-MAR-1998.

PD 16-SEP-1996; 96CA-02185605.

PF 16-SEP-1996; 96CA-02185605.

PR (MAUK/) MAUK A G.

PA (WANL/) WAN L.

PA (LEEH/) LEE H.

PA (BRAY/) BRAYER G D.

PA (TONG/) TONG H.

PA (SMIT/) SMITH M.

PI Wan L, Mauk AG, Lee H, Brayer GD, Tong H, Smith M;

XX WPI; 1998-388655/34.

XX New modified myoglobin with amino acid alterations has increased

PT peroxidase activity - used for catalysing oxidation of substrates with

PT peroxides, e.g. in waste treatment and as label for antigens.

XX Claim 21; Page; 23pp; English.

CC Sequences AAW62270-W62273, and AAW29739-W29744, are modified synthetic

CC horse heart myoglobin (Mb) proteins, produced by random mutagenesis of

CC the synthetic Mb gene. This present sequence contains a substitution at

CC Lys45Glu, which has been found to enhance the peroxidase activity of the

CC protein in general, and it may also exhibit enhanced Mn(II) peroxidase

CC activity as well. This is due to the enhanced rate of reaction of Fe(III)

CC Mb with hydrogen peroxide and also because the substitution at residue 45
CC creates a new Mn binding site which is stronger than the natural Mn
CC binding site of Mb. This sequence also contains substitutions at
CC Thr39Ile, Phe46Leu, and Ile107Phe, which have been shown to enhance the
CC peroxidase activity of Mb. These protein can be used to catalyse oxidation
CC of a substrate by peroxidase, e.g. oxidation of Mn2+. Peroxidases are
CC used for oxidising phenolics to polymers, as labels for e.g. antigens,
CC antibodies or oligonucleotides, in diagnostic assays and for oxidative
CC treatment of waste streams. NB. This sequence is not given in the
CC specification but was created from the wild type sequence given by the
CC inventors. The positions of the substitutions given do not include the
CC initiation codon
XX
SQ Sequence 154 AA;

Query Match 88.0%; Score 699; DB 2; Length 154;
Best Local Similarity 86.2%; Pred. No. 1.2e-69;
Matches 131; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

QY 2 LSEGEQVLVHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASEDL 61
Db 3 LSDGEWQVNLVWVGKVEADIAHGQEVLRFTGHPELTLEKFDKFKHLKTEAEKASEDL 62

QY 62 KKHGTVLVTALGAILKKKGHEALKPLAQSHATKHKIPKYLEFTSEALIIHVLHSHRHPG 121

Db 63 KKHGTVLVTALGAILKKKGHEALKPLAQSHATKHKIPKYLEFTSEALIIHVLHSHRHPG 122

QY 122 DFGADAQAGAMTKALELFRNDIAAKYKELGYQG 153

Db 123 DFGADAQAGAMTKALELFRNDIAAKYKELGYQG 154

RESULT 13

ADC31106

ID ADC31106 standard; protein; 166 AA.

XX ADC31106;

XX 18-DEC-2003 (first entry)

XX Human novel polypeptide sequence, SEQ ID NO:1188.

XX Human; diagnostic; drug screening; forensics; gene mapping;

KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;

KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;

KW ulcers; osteoporosis; autoimmune disease; cancer;

KW molecular weight marker; food supplement; antiparkinsonian; nootropic;

KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;

KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;

KW gene therapy; chromosome 22q13.1.

XX Homo sapiens.

OS WO2003029271-A2.

PN 10-APR-2003.

PD 24-SEP-2002; 2002WO-US030474.

PF 24-SEP-2001; 2001US-0324631P.

PR (HYSE-) HYSEQ INC.

XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;

XX Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;

XX Haley-Vicente D, Drmanac RT;

XX WPI; 2003-371981/35.

DR N-FSDB; ADC30135.

XX New polynucleotide and polypeptide useful for diagnosing, preventing or

XX treating conditions such as neurodegenerative diseases, anemias, platelet

XX disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or

Db 115 KKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIQVLQSKHPG 174
QY 122 DFGADAQGAMNKALELFRKDIAAKYKELGYQG 153
Db 175 DFGADAQGAMNKALELFRKDIAAKYKELGYQG 206
RESULT 15
AAW62273
ID AAW62273 standard; protein; 154 AA.
XX
AC AAW62273;
XX
DT 26-OCT-1998 (first entry)
XX
DE Modified myoglobin protein 4.
XX
KW Heart; myoglobin; Mb; random mutagenesis; peroxidase activity; Fe(III);
KW Mn(II) peroxidase activity; hydrogen peroxide; Mn binding site; catalyze;
KW phenolic; polymer; antigen; antibody; diagnostic assay.
XX
OS Equus sp.
OS Synthetic.
XX
PN CA2185605-A.
XX
PD 17-MAR-1998.
XX
PF 16-SEP-1996; 96CA-02185605.
XX
PR 16-SEP-1996; 96CA-02185605.
XX
PA (MAUK/) MAUK A G.
PA (WANL/) WAN L.
PA (LEEH/) LEE H.
PA (BRAY/) BRAYER G D.
PA (TONG/) TONG H.
PA (SMIT/) SMITH M.
XX
PI Wan L, Mauk AG, Lee H, Brayer GD, Tong H, Smith M;
XX WPI; 1998-388655/34.
DR
XX
PT New modified myoglobin with amino acid alterations has increased
PT peroxidase activity - used for catalysing oxidation of substrates with
PT peroxides, e.g. in waste treatment and as label for antigens.
XX
PS Claim 23; Page; 23pp; English.
XX
CC Sequences AAW62270-W62273, and AAW29739-W29744, are modified synthetic
CC horse heart myoglobin (Mb) proteins, produced by random mutagenesis of
CC the synthetic Mb gene. This present sequence contains a substitution at
CC Lys45Glu, which has been found to enhance the peroxidase activity of the
CC protein in general, and it may also exhibit enhanced Mn(II) peroxidase
CC activity as well. This is due to the enhanced rate of reaction of Fe(III)
CC Mb with hydrogen peroxide and also because the substitution at residue 45
CC creates a new Mn binding site which is stronger than the natural Mn
CC binding site of Mb. This protein also contains a substitution at
CC Lys83Glu, which was also found to be required for Mn binding, and
CC substitutions at His97Leu and Ser92Ala to increase the rate of Mn(II)
CC turnover. These protein can be used to catalyze oxidation of a substrate
CC by peroxidase, e.g. oxidation of Mn2+. Peroxidases are used for oxidising
CC phenolics to polymers, as labels for e.g. antigens, antibodies or
CC oligonucleotides, in diagnostic assays and for oxidative treatment of
CC waste streams. NB. This sequence is not given in the specification but
CC was created from the wild type sequence given by the inventors. The
CC positions of the substitutions given do not include the initiation codon
XX Sequence 154 AA;

Query Match 87.8%; Score 697; DB 2; Length 154;
Best Local Similarity 86.2%; Pred. No. 2e-69;
Matches 131; Conservative 10; Mismatches 11; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHGQDILIRLFXSHDETLEKFDPRFKHLKTEAEMKASEDL 61
Db 3 LSDGEWQQVLNVWGKVEADIAHGQEVLRIRLFTGHPETLEKFDPRFKHLKTEAEMKASEDL 62
QY 62 KKHGVTVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISEATIHVLSRHPG 121
Db 63 KEHGTVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISDAIHVLSRHPG 122
QY 122 DFGADAQGAMNKALELFRKDIAAKYKELGYQG 153
Db 123 DFGADAQGAMTKALELFRNDIAAKYKELGYQG 154

Search completed: August 10, 2004, 15:26:14
Job time : 27.4879 secs

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:23:41 ; Search time 7.77966 Seconds
(without alignments)
1015.311 Million cell updates/sec

Title: US-09-455-978B-76
Perfect score: 794
Sequence: 1 VLSEGEWQLVHVWAKVEAD.....ALELFRKDIAAKYKELGYOG 153

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/prodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	794	100.0	153	3	US-08-992-176-7
2	778.5	98.0	152	3	US-08-992-176-10
3	717	90.3	154	3	US-08-710-330A-11
4	696	87.7	154	3	US-08-817-787-18
5	584	73.6	113	2	US-08-710-330A-2
6	525	66.1	113	2	US-08-710-330A-1
7	515	64.9	113	2	US-08-710-330A-5
8	505	63.6	113	2	US-08-710-330A-4
9	504	63.5	113	2	US-08-710-330A-3
10	240	30.2	60	3	US-08-817-787-22
11	240	30.2	64	3	US-08-817-787-24
12	213	26.8	57	3	US-08-817-787-20
13	162	20.4	141	1	US-08-240-712-18
14	162	20.4	141	1	US-08-443-890-18
15	162	20.4	141	2	US-08-484-686B-65
16	162	20.4	141	3	US-08-463-160B-66
17	162	20.4	141	5	PT-US92-09752-18
18	152	19.1	141	3	US-09-058-562-18
19	150	18.9	146	4	US-09-353-719-1
20	140	17.6	141	2	US-08-627-173-21
21	140	17.6	141	2	US-08-535-882A-21
22	140	17.6	141	3	US-08-316-424A-7
23	140	17.6	141	3	US-09-005-546-21
24	140	17.6	141	4	US-08-477-669-7
25	140	17.6	141	4	US-10-128-581-26
26	135	17.0	146	2	US-08-619-708A-4
27	128	16.1	141	2	US-08-627-173-19

28	128	16.1	141	2	US-08-535-882A-19	Sequence 19, Appl
29	128	16.1	141	3	US-08-316-424A-5	Sequence 5, Appl1
30	128	16.1	141	3	US-09-005-546-19	Sequence 19, Appl
31	128	16.1	141	4	US-08-477-669-5	Sequence 5, Appl1
32	128	16.1	141	4	US-10-128-581-24	Sequence 24, Appl
33	127	16.0	146	2	US-08-627-173-22	Sequence 22, Appl
34	127	16.0	146	2	US-08-535-882A-22	Sequence 22, Appl
35	127	16.0	146	3	US-08-316-424A-8	Sequence 8, Appl1
36	127	16.0	146	3	US-09-005-546-22	Sequence 22, Appl
37	127	16.0	146	4	US-08-477-669-8	Sequence 8, Appl1
38	127	16.0	146	4	US-10-128-581-27	Sequence 27, Appl
39	125	15.7	146	1	US-08-240-712-24	Sequence 24, Appl
40	125	15.7	146	1	US-08-443-890-24	Sequence 24, Appl
41	125	15.7	146	3	US-09-058-562-22	Sequence 22, Appl
42	125	15.6	146	5	PT-US92-09752-24	Sequence 24, Appl
43	124	15.6	141	2	US-08-619-708A-2	Sequence 2, Appl1
44	124	15.6	146	1	US-08-240-712-21	Sequence 21, Appl
45	124	15.6	146	1	US-08-443-890-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-08-992-176-7
; Sequence 7, Application US/08992176
; Patent No. 6125331
; GENERAL INFORMATION:
; APPLICANT: TOH, Hiroyuki
; TITLE OF INVENTION: STRUCTURAL ALIGNMENT METHOD MAKING USE OF A DOUBLE
; FILE REFERENCE: 9200-0001-2
; CURRENT APPLICATION NUMBER: US/08/992,176
; EARLIER FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: JP 8-340727
; EARLIER FILING DATE: 1996-12-20
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:alpha-protein
; OTHER INFORMATION: (Imbc)
US-08-992-176-7

Query Match	100.0%	Score 794;	DB 3;	Length 153;
Best Local Similarity	100.0%;	Pred. No. 2e-81;	Mismatches 0;	Indels 0; Gaps 0;
Matches 153;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	VLSEGEWQLVHVWAKVEADVAGHQDILIRLFKSHPETLEKFRFKHLKTEAEKASD	60	
Db	1	VLSEGEWQLVHVWAKVEADVAGHQDILIRLFKSHPETLEKFRFKHLKTEAEKASD	60	
Qy	61	LKKHGVTVLTALGAILKKKGHEAEKPLAQSHATKHKIPIKYLEFISEAIIHVLSRHP	120	
Db	61	LKKHGVTVLTALGAILKKKGHEAEKPLAQSHATKHKIPIKYLEFISEAIIHVLSRHP	120	
Qy	121	GDFGADAQAMNKALELFRKDIAAKYKELGYOG	153	
Db	121	GDFGADAQAMNKALELFRKDIAAKYKELGYOG	153	

RESULT 2
US-08-992-176-10
; Sequence 10, Application US/08992176
; Patent No. 6125331
; GENERAL INFORMATION:
; APPLICANT: TOH, Hiroyuki
; TITLE OF INVENTION: STRUCTURAL ALIGNMENT METHOD MAKING USE OF A DOUBLE
; FILE REFERENCE: 9200-0001-2

```
;
; CURRENT APPLICATION NUMBER: US/08/992,176
; CURRENT FILING DATE: 1997-12-17
; EARLIER APPLICATION NUMBER: JP 8-340727
; EARLIER FILING DATE: 1996-12-20
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 10
; LENGTH: 152
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: alpha-protein
; OTHER INFORMATION: (lmhc)
US-08-992-176-10

Query Match          98.0%; Score 778.5; DB 3; Length 152;
Best Local Similarity 99.3%; Pred. No. 1.1e-79;
Matches 152; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 VLSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEKASEDL 60
Db 1 VLSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEKASEDL 60
QY 61 LKKGVTVLTAALGAILKKKGHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
Db 61 LKKGVTVLTAALGAILKKKGHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHS-HP 119
QY 121 GDFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
Db 120 GDFGADAQAGAMNKALELFRKDIAAKYKELGYQG 152

RESULT 3
US-08-710-330A-11
; Sequence 11, Application US/08710330A
; Patent No. 5854041
; GENERAL INFORMATION:
; APPLICANT: Brayer, Gary D.
; APPLICANT: Lee, Hung
; APPLICANT: Mauk, Grant A.
; APPLICANT: Smith, Michael
; APPLICANT: Tong, Harry
; APPLICANT: Wan, Lianglu
; TITLE OF INVENTION: MYOGLOBIN WITH PEROXIDASE ACTIVITY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,330A
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: GAY, David A.
; REGISTRATION NUMBER: 39,200
; REFERENCE/DOCKET NUMBER: P-SM 2262
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 154 amino acids
; TYPE: amino acid
; STRANDEDNESS:
```

```
;
; TOPOLOGY: linear
US-08-710-330A-11

Query Match          90.3%; Score 717; DB 2; Length 154;
Best Local Similarity 88.2%; Pred. No. 9e-73;
Matches 134; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEKASEDL 61
Db 3 LSDGEWQQLVNWGVKVEADIAHGQGVLRFTGHPETLEKDFKFKHLKTEAEKASEDL 62
QY 62 KKHGVTVLTAALGAILKKKGHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPG 121
Db 63 KKHGVTVLTAALGAILKKKGHEAEKPLAQSHATKHKIPKYLEFISDAIHHVLSKHGPG 122

QY 122 DFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
Db 123 DFGADAQAGAMNKALELFRNDIAAKYKELGFGQ 154

RESULT 4
US-08-817-787-18
; Sequence 18, Application US/08817787
; Patent No. 6294353
; GENERAL INFORMATION:
; APPLICANT: Lupas, Andrei
; APPLICANT: Pack, Peter
; TITLE OF INVENTION: TARGETED HETERO-ASSOCIATION OF
; TITLE OF INVENTION: RECOMBINANT PROTEINS TO MULTI-FUNCTIONAL COMPLEXES
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,787
; FILING DATE: 23-SEP-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04117
; FILING DATE:
; PRIOR APPLICATION DATA: EP 94 11 6558.1
; FILING DATE: 20-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MORPHO/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 154 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-08-817-787-18

Query Match          87.7%; Score 696; DB 3; Length 154;
Best Local Similarity 84.2%; Pred. No. 2e-70;
Matches 128; Conservative 14; Mismatches 10; Indels 0; Gaps 0;
```

QY 2 LSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKFHLKTEAEKASEDL 61
 Db 3 LSGEGWQLVNVWGVKVEADIPRGQGVLRIRLFKGHPETLEKDFRKFHLKSEDEMKASEDL 62
 QY 62 KKHGVTVLTAIGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISEAIIHVLHSHRHPG 121
 Db 63 KKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIQVLSKHG 122
 QY 122 DFGADAGGAMKALEFRKDIAAKYKELGVQ 153
 Db 123 DFGADAGGAMKALEFRKDMASNYKELGVQ 154

RESULT 5

US-08-710-330A-2
 ; Sequence 2, Application US/08710330A
 ; Patent No. 5854041
 ; GENERAL INFORMATION:
 ; APPLICANT: Brayer, Gary D.
 ; APPLICANT: Lee, Hung
 ; APPLICANT: Mauk, Grant A.
 ; APPLICANT: Smith, Michael
 ; APPLICANT: Tong, Harry
 ; APPLICANT: Wan, Lianglu
 ; TITLE OF INVENTION: MYOGLOBIN WITH PEROXIDASE ACTIVITY
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Campbell & Flores LLP
 ; STREET: 4370 La Jolla Village Drive, Suite 700
 ; CITY: San Diego
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 92122

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/710,330A
 FILING DATE: 16-SEP-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: GAY, David A.
 REGISTRATION NUMBER: 39,200
 REFERENCE/DOCKET NUMBER: P-SM 2262
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 113 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear

US-08-710-330A-2

Query Match 73.6%; Score 584; DB 2; Length 113;
 Best Local Similarity 100.0%; Pred. No. 5e-58;
 Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKFHLKTEAEKASED 60
 Db 1 VLSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKFHLKTEAEKASED 60
 QY 61 LKHGVTVLTAIGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISEAIIH 113
 Db 61 LKHGVTVLTAIGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISEAIIH 113

RESULT 6

US-08-710-330A-1

; Sequence 1, Application US/08710330A
 ; Patent No. 5854041
 ; GENERAL INFORMATION:
 ; APPLICANT: Brayer, Gary D.
 ; APPLICANT: Lee, Hung
 ; APPLICANT: Mauk, Grant A.
 ; APPLICANT: Smith, Michael
 ; APPLICANT: Tong, Harry
 ; APPLICANT: Wan, Lianglu
 ; TITLE OF INVENTION: MYOGLOBIN WITH PEROXIDASE ACTIVITY
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Campbell & Flores LLP
 ; STREET: 4370 La Jolla Village Drive, Suite 700
 ; CITY: San Diego
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 92122

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/710,330A
 FILING DATE: 16-SEP-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: GAY, David A.
 REGISTRATION NUMBER: 39,200
 REFERENCE/DOCKET NUMBER: P-SM 2262
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 113 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 US-08-710-330A-1

Query Match 66.1%; Score 525; DB 2; Length 113;
 Best Local Similarity 87.5%; Pred. No. 2.1e-51;
 Matches 98; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKFHLKTEAEKASEDL 61
 Db 2 LSGEGWQLVNVWGVKVEADIPRGQGVLRIRLFKGHPETLEKDFRKFHLKSEDEMKASEDL 61
 QY 62 KKHGVTVLTAIGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISEAIIH 113
 Db 62 KKHGVTVLTAIGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISEAIIH 113

RESULT 7

US-08-710-330A-5
 ; Sequence 5, Application US/08710330A
 ; Patent No. 5854041
 ; GENERAL INFORMATION:
 ; APPLICANT: Brayer, Gary D.
 ; APPLICANT: Lee, Hung
 ; APPLICANT: Mauk, Grant A.
 ; APPLICANT: Smith, Michael
 ; APPLICANT: Tong, Harry
 ; APPLICANT: Wan, Lianglu
 ; TITLE OF INVENTION: MYOGLOBIN WITH PEROXIDASE ACTIVITY
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Campbell & Flores LLP
 ; STREET: 4370 La Jolla Village Drive, Suite 700
 ; CITY: San Diego
 ; STATE: California

```

; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,330A
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: GAY, David A.
; REGISTRATION NUMBER: 39,200
; REFERENCE/DOCKET NUMBER: P-SM 2262
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 113 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-710-330A-5
Query Match 64.9%; Score 515; DB 2; Length 113;
Best Local Similarity 85.6%; Pred. No. 2.8e-50;
Matches 95; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

QY 2 LSGGEWQLVHVWAKVEADVAGHGQDILIRLPKSHPETLEKDFRFXHLKTEAEMKASEDL 61
DB 2 LSGGEWQLVNWGKVEADIPGQGQEVLRIRLFGHPETLEKDFRFXHLKSEDEMKASEDL 61

QY 62 KKHGVTVLTALGAILKKKGHEAEKPLAQSHATKHKIPVKYLEFISEAII 112
DB 62 KKHGATVLTALGGILKKKGHEAEIKPLAQSHATKHKIPVKYLEFISECII 112

RESULT 8
US-08-710-330A-4
; Sequence 4, Application US/08710330A
; Patent No. 5854041
; GENERAL INFORMATION:
; APPLICANT: Brayer, Gary D.
; APPLICANT: Lee, Hung
; APPLICANT: Mauk, Grant A.
; APPLICANT: Smith, Michael
; APPLICANT: Tong, Harry
; APPLICANT: Wan, Lianglu
; TITLE OF INVENTION: MYOGLOBIN WITH PEROXIDASE ACTIVITY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,330A
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: GAY, David A.
; REGISTRATION NUMBER: 39,200
; REFERENCE/DOCKET NUMBER: P-SM 2262
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 113 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-710-330A-3
Query Match 64.9%; Score 515; DB 2; Length 113;
Best Local Similarity 85.6%; Pred. No. 2.8e-50;
Matches 95; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

QY 2 LSGGEWQLVHVWAKVEADVAGHGQDILIRLPKSHPETLEKDFRFXHLKTEAEMKASEDL 61
DB 2 LSGGEWQLVNWGKVEADIPGQGQEVLRIRLFGHPETLEKDFRFXHLKSEDEMKASEDL 61

QY 62 KKHGVTVLTALGAILKKKGHEAEKPLAQSHATKHKIPVKYLEFISEAII 112
DB 62 KKHGATVLTALGGILKKKGHEAEIKPLAQSHATKHKIPVKYLEFISECII 112

RESULT 9
US-08-710-330A-3
; Sequence 3, Application US/08710330A
; Patent No. 5854041
; GENERAL INFORMATION:
; APPLICANT: Brayer, Gary D.
; APPLICANT: Lee, Hung
; APPLICANT: Mauk, Grant A.
; APPLICANT: Smith, Michael
; APPLICANT: Tong, Harry
; APPLICANT: Wan, Lianglu
; TITLE OF INVENTION: MYOGLOBIN WITH PEROXIDASE ACTIVITY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,330A
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: GAY, David A.
; REGISTRATION NUMBER: 39,200
; REFERENCE/DOCKET NUMBER: P-SM 2262
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 113 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-710-330A-3
Query Match 63.5%; Score 504; DB 2; Length 113;
Best Local Similarity 83.9%; Pred. No. 4.8e-49;
Matches 94; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

QY 2 LSGGEWQLVHVWAKVEADVAGHGQDILIRLPKSHPETLEKDFRFXHLKTEAEMKASEDL 61
DB 2 LSGGEWQLVNWGKVEADIPGQGQEVLRIRLFGHPETLEKDFRFXHLKSEDEMKASEDL 61
```

```

; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 113 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-710-330A-4
Query Match 63.6%; Score 505; DB 2; Length 113;
Best Local Similarity 83.0%; Pred. No. 3.7e-49;
Matches 93; Conservative 11; Mismatches 8; Indels 0; Gaps 0;

QY 2 LSGGEWQLVHVWAKVEADVAGHGQDILIRLPKSHPETLEKDFRFXHLKTEAEMKASEDL 61
DB 2 LSGGEWQLVNWGKVEADVAGHGQEVLRIRLFGHPETLEKDFRFXHLKTEAEMKASEDL 61

QY 62 KKHGVTVLTALGAILKKKGHEAEKPLAQSHATKHKIPVKYLEFISEAIIH 113
DB 62 KKHGNTVLTALGGILKKKGHEAEVGHIAESHANKHKIPVKYLEFISDAIIH 113

RESULT 9
US-08-710-330A-3
; Sequence 3, Application US/08710330A
; Patent No. 5854041
; GENERAL INFORMATION:
; APPLICANT: Brayer, Gary D.
; APPLICANT: Lee, Hung
; APPLICANT: Mauk, Grant A.
; APPLICANT: Smith, Michael
; APPLICANT: Tong, Harry
; APPLICANT: Wan, Lianglu
; TITLE OF INVENTION: MYOGLOBIN WITH PEROXIDASE ACTIVITY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,330A
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: GAY, David A.
; REGISTRATION NUMBER: 39,200
; REFERENCE/DOCKET NUMBER: P-SM 2262
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 113 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-710-330A-3
Query Match 63.5%; Score 504; DB 2; Length 113;
Best Local Similarity 83.9%; Pred. No. 4.8e-49;
Matches 94; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

QY 2 LSGGEWQLVHVWAKVEADVAGHGQDILIRLPKSHPETLEKDFRFXHLKTEAEMKASEDL 61
DB 2 LSGGEWQLVNWGKVEADIPGQGQEVLRIRLFGHPETLEKDFRFXHLKSEDEMKASEDL 61
```

```

Db      2  LSDGEQAVLNAMKQVADVAGHQGVILRLFTGHDPETLEKFDKFKHLKTEAMKASEDL 6
Qy      62  KKHGVTVLTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIH 113
Db      62  KKHGTVLTALGGILKKKGHHEAEVKHLAESHANKHKVPKYLEFISDAIIH 113

RESULT 10
US-08-817-787-22
; Sequence 22, Application US/08817787
; Patent No. 6294353
; GENERAL INFORMATION:
; APPLICANT: Pack, Peter
; APPLICANT: Lupas, Andrei
; TITLE OF INVENTION: TARGETED HETERO-ASSOCIATION OF
; TITLE OF INVENTION: RECOMBINANT PROTEINS TO MULTI-FUNCTIONAL COMPLEXES
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,787
; FILING DATE: 23-SEP-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04117
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94 11 6558.1
; FILING DATE: 20-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MORPHO/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 60 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-817-787-22

Query Match 30.2%; Score 240; DB 3; Length 60;
Best Local Similarity 90.2%; Pred.No. 8.6e-20;
Matches 46; Conservative 2; Mismatches 3; Indels 0; Gaps 0

QY      48  HLKTEAMKASEDLKKKHGVTVLTALGAILKKKGHHEAELKPLAQSHATKHK 98
Db      3  HLKSEDMKASEDLKKHGATVLTALGGILKKKGHHEAELKPLAQSHATKHK 53

RESULT 11
US-08-817-787-24
; Sequence 24, Application US/08817787
; Patent No. 6294353
; GENERAL INFORMATION:
; APPLICANT: Pack, Peter
; APPLICANT: Lupas, Andrei
; TITLE OF INVENTION: TARGETED HETERO-ASSOCIATION OF
; TITLE OF INVENTION: RECOMBINANT PROTEINS TO MULTI-FUNCTIONAL COMPLEXES
; NUMBER OF SEQUENCES: 36

```

APPLICATION NUMBER: PCT/EP95/04117
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94 11 6558.1
FILING DATE: 20-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: MORPHO/1
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 57 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-817-787-20

Query Match 26.8%; Score 213; DB 3; Length 57;
Best Local Similarity 80.4%; Pred. No. 8.6e-17;
Matches 37; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHWAKVEADVAGHGODILIRLFKSHPETLEKFDREK 47
DB 5 LSDGEWQLVNWGVKVEADIPGRGQEVLRILFKGHPETLEKFDKFK 50

RESULT 13

US-08-240-712-18
Sequence 18, Application US/08240712
Patent No. 5599907
GENERAL INFORMATION:
APPLICANT: ANDERSON, DAVID C.
APPLICANT: MATHEWS, ANTONY JAMES
APPLICANT: STETLER, GARY L.
TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC
TITLE OF INVENTION: HEMOGLOBINS
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/240,712
FILING DATE: 09-MAY-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/09752
FILING DATE: 13-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: COOPER, IVER P
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: ANDERSON=6
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 141 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: peptide
US-08-240-712-18

Query Match 20.4%; Score 162; DB 1; Length 141;
Best Local Similarity 28.1%; Pred. No. 1.6e-10;
Matches 41; Conservative 29; Mismatches 70; Indels 6; Gaps 1;
QY 2 LSEGEWQLVHWAKVEADVAGHGODILIRLFKSHPETLEKFDREKFKLTKTEAEKASEDL 61
DB 2 LTKTERTIIVSWAKISTOADTIGTETLERLFLSHPTKTYFPHFDLHFGSAQLRA--- 57
QY 62 KKHGVTVLTAIGAILKKKGHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHRRHPG 121
DB 58 --HGSKVVAAGDAVKSIDDIGALSSELHAYILRVDVNVNFKLSHCLLVTLAARFFA 115
QY 122 DFGADAQAGMNALELFRKDIKAYK 147
DB 116 DFTAEAAAWDKFLSVSVSVLTKYR 141

RESULT 14

US-08-443-890-18
Sequence 18, Application US/08443890
Patent No. 5739011
GENERAL INFORMATION:
APPLICANT: ANDERSON, DAVID C.
APPLICANT: MATHEWS, ANTONY JAMES
APPLICANT: STETLER, GARY L.
TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC
TITLE OF INVENTION: HEMOGLOBINS
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/443,890
FILING DATE: 31-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/240,712
FILING DATE: 09-MAY-1994
APPLICATION NUMBER: PCT/US92/09752
FILING DATE: 13-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: COOPER, IVER P
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: ANDERSON=6
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 141 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-443-890-18

Query Match 20.4%; Score 162; DB 1; Length 141;
Best Local Similarity 28.1%; Pred. No. 1.6e-10;
Matches 41; Conservative 29; Mismatches 70; Indels 6; Gaps 1;

QY	2	LSGSEQLVLHWAKVEADVAGHQDIIIRLFKSHPETLEKFDPRFKHLKTEAEMKASEDL	61
		2 LKERTIIIVSWWAKISTQADTIGTETLERLFLSHPEQTKYFPFDFLHPGSAQLRA---	57
QY	62	KKHGVTVLPALGAILKKKGHHAEALPKLAQGHATKHKIKYLFESIALIHWLHGRHPG	121
Db	58	--HGSKVAAVAGDAVKSIDDIGALSKLSLSELHAVILRDPVNFKLGHCLLVTLARFPA	115
QY	122	DFGADAGQAMNKALELFRKDIAAAYK	147
Db	116	DFTAEAAHAWDKFLSVSWSVLTKYR	141

RESULT 15

US-08-484-686B-65
; Sequence 65, Application US/08484686B
; Patent No. 5827693
; GENERAL INFORMATION:
; APPLICANT: De Angelo, Joseph
; APPLICANT: Motwani, Nalini
; APPLICANT: Bajwa, Wajeeh
; TITLE OF INVENTION: Expression of Recombinant Hemoglobin and
; TITLE OF INVENTION: Hemoglobin Variants in Yeast
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennle & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,686B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/368,407
; FILING DATE: 29-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/876,290
; FILING DATE: 29-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/684,611
; FILING DATE: 12/APR/1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Flintoft, Gerald J.
; REGISTRATION NUMBER: 20,823
; REFERENCE/DOCKET NUMBER: 6666-043-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-484-686B-65

```

Query Match      20.4%; Score 162; DB 2; Length 141;
Best Local Similarity 28.1%; Pres. No. 1.6e-10;
Matches 41; Conservative 25; Mismatches 70; Indels 6; Gaps 1;

Qy      2 LSEGEVLVHWAKVEADVAGHGQDILRLFKSHPELLEKEDRFKHLKTEAEKMSDDL 61
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2 LTKTERIIIVSMMAKLSQTADTGTETLERLSHPQTYTFPHFDLPFGSAQLRA---- 57

```

Qy 62 KXHGVTVLTAIGALKKKKGHHEAEKPLAQSHATKHKIPKYLEPISEAIHVLHSRHPG 121
 Db 58 --HGSKVVAAGVDAAVKSIDDITGGALSKEISELHAYILRVDPVNFKLLSHCLLVTLAARFPA 115
 Qy 122 DFGADAQAGAMNKALELFRKDTAAKYK 147
 Db 116 DFTAEAAHAAWDKFLSVSSVLTEKYR 141

Search completed: August 10, 2004, 15:30:20
Job time : 8.77966 secs

Blank

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:28:46 ; Search time 22.2276 Seconds
(without alignments)
2159.184 Million cell updates/sec

Title: US-09-455-978b-76
Perfect score: 794
Sequence: 1 VLSEGEQLVHLVWAKVEAD.....ALELFRKDIAAKYKELGYQG 153

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/ECTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	696	87.7	154	16	US-10-408-765A-106	Sequence 106, Appl
2	162	20.4	141	9	US-09-977-577-19	Sequence 19, Appl
3	152	19.1	31	15	US-10-289-009-17	Sequence 17, Appl
4	140	17.6	141	9	US-09-839-164-7	Sequence 7, Appl
5	140	17.6	141	14	US-10-128-581-26	Sequence 26, Appl
6	140	17.6	141	16	US-10-463-699-28	Sequence 28, Appl
7	133	16.8	31	15	US-10-289-009-4	Sequence 4, Appl
8	133	16.8	154	14	US-10-378-029-102	Sequence 102, Appl
9	132	16.6	142	12	US-10-424-599-227247	Sequence 227247,
10	132	16.6	142	14	US-10-205-219-189	Sequence 189, Appl
11	128	16.1	141	9	US-09-839-164-5	Sequence 5, Appl
12	128	16.1	141	14	US-10-128-581-24	Sequence 24, Appl
13	128	16.1	141	16	US-10-463-699-26	Sequence 26, Appl
14	127	16.0	146	9	US-09-839-164-8	Sequence 8, Appl
15	127	16.0	146	14	US-10-128-581-27	Sequence 27, Appl

16	127	16.0	146	16	US-10-463-699-29	Sequence 29, Appl
17	125	15.7	146	9	US-09-977-577-20	Sequence 20, Appl
18	124	15.6	146	9	US-09-977-577-17	Sequence 17, Appl
19	120	15.1	146	9	US-09-839-164-6	Sequence 6, Appl
20	120	15.1	146	14	US-10-128-581-25	Sequence 25, Appl
21	120	15.1	146	16	US-10-463-699-27	Sequence 27, Appl
22	114	14.4	147	9	US-09-147-490-8	Sequence 8, Appl
23	112	14.1	141	9	US-09-839-164-2	Sequence 2, Appl
24	112	14.1	141	9	US-09-977-577-14	Sequence 14, Appl
25	112	14.1	141	9	US-09-977-577-18	Sequence 18, Appl
26	112	14.1	141	12	US-10-280-725B-8	Sequence 8, Appl
27	112	14.1	141	13	US-10-085-853-31	Sequence 31, Appl
28	112	14.1	141	14	US-10-128-581-21	Sequence 21, Appl
29	112	14.1	141	14	US-10-280-679B-8	Sequence 8, Appl
30	112	14.1	141	16	US-10-463-699-23	Sequence 23, Appl
31	112	14.1	142	15	US-10-435-666-7	Sequence 7, Appl
32	111	14.0	141	16	US-10-408-765A-310	Sequence 310, Appl
33	104	13.1	122	15	US-10-115-482-52	Sequence 52, Appl
34	103	13.0	146	9	US-09-839-164-4	Sequence 4, Appl
35	103	13.0	146	9	US-09-977-577-15	Sequence 15, Appl
36	103	13.0	146	13	US-10-085-853-33	Sequence 33, Appl
37	103	13.0	146	14	US-10-128-581-23	Sequence 23, Appl
38	103	13.0	146	16	US-10-463-699-25	Sequence 25, Appl
39	103	13.0	146	16	US-10-408-765A-3019	Sequence 3019, Appl
40	101	12.7	19	15	US-10-289-009-12	Sequence 12, Appl
41	101	12.7	20	12	US-10-103-395-189	Sequence 189, Appl
42	100	12.6	147	9	US-09-977-577-21	Sequence 21, Appl
43	99	12.5	146	9	US-09-977-577-16	Sequence 16, Appl
44	98	12.3	146	16	US-10-408-765A-84	Sequence 84, Appl
45	96	12.1	18	15	US-10-289-009-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1

US-10-408-765A-106
; Sequence 106, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 106
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-106

Query Match 87.7%; Score 696; DB 16; Length 154;
Best Local Similarity 84.2%; Pred. No. 2.3e-69;
Matches 128; Conservative 14; Mismatches 10; Indels 0; Gaps 0;

Qy	2	LSGEQQLVHLVWAKVEADVAGHGDILIRLFKSHPETLEKFDPRFKHLKTEAEKASDGL 61
Db	3	LSDGEQLVNLVWGVKVEADIPGHGQEVILIRLFKSHPETLEKFDPRFKHLKSEDEKASDGL 62
Qy	62	KHGVTVTLTGAILKKKGHHEAEIKPLAQSHATKHGKTIPIKYLEFISEAIIHVLHSRHPG 121
Db	63	KHGVTVTLTGAILKKKGHHEAEIKPLAQSHATKHGKTIPIKYLEFISEAIIHVLHSRHPG 122
Qy	122	DFGADAQAGMNALELFRKDIAAKYKELGYQG 153


```
RESULT 5
US-10-581-26
; Sequence 26, Application US/10128581
; Publication No. US20030104984A1
; GENERAL INFORMATION:
; APPLICANT: TSRYLOVA, IRENA
; WOLPE, STEPHEN D.
; TITLE OF INVENTION: HEMOGLOBIN ALPHA CHAIN PEPTIDE FRAGMENTS USEFUL
; FOR INHIBITING STEM CELL PROLIFERATION
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 MB disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/128,581
; FILING DATE: 24-APR-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/617,840
; FILING DATE: 17-JUL-2000
; APPLICATION NUMBER: US 09/005,546
; FILING DATE: 12-JAN-1998
; APPLICATION NUMBER: US 08/535,882
; FILING DATE: 28-SEP-1995
; APPLICATION NUMBER: US 08/316,424
; FILING DATE: 30-SEP-1994
; APPLICATION NUMBER: PCT/US94/03349
; FILING DATE: 29-MAR-1994
; APPLICATION NUMBER: US 08/040,942
; FILING DATE: 31-MAR-1993
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-10-581-26
Query Match 17.6%; Score 140; DB 14; Length 141;
Best Local Similarity 27.6%; Pred. No. 2.1e-07;
Matches 42; Conservative 20; Mismatches 74; Indels 16; Gaps 4;
QY 1 VLSGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKPDFR--KHLKTEAEKAS 58
Db 1 VLSAADKANVKAAGKVGQAGAHGAELERMFGLGPTTKTYFPHNLSH-----GS 52
QY 59 EDLKKHGVTVTALGAILKKKGHHE---AELKPLAQSHATKHKIPKYLEFISEAIHVL 115
Db 53 DOVKAGQKVADAL---TKAVGHLDLPGALSALSDLHAHLKLRVDPVNFKLLSHCLLVTL 109
QY 116 HSRHPDGDGADAQAGAMNKALELFRKDIAAKYK 147
Db 110 AAHPDDPNPSVHASLDKFLANVSTVLTSKYR 141
RESULT 6
US-10-463-699-28
; Sequence 28, Application US/10463699
; Publication No. US20040081640A1
; GENERAL INFORMATION:
; APPLICANT: TSRYLOVA, IRENA
; WOLPE, STEPHEN D.
; TITLE OF INVENTION: HEMOGLOBIN ALPHA CHAIN PEPTIDE FRAGMENTS USEFUL
; FOR INHIBITING STEM CELL PROLIFERATION
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 MB disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/128,581
; FILING DATE: 24-APR-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/617,840
; FILING DATE: 17-JUL-2000
; APPLICATION NUMBER: US 09/005,546
; FILING DATE: 12-JAN-1998
; APPLICATION NUMBER: US 08/535,882
; FILING DATE: 28-SEP-1995
; APPLICATION NUMBER: US 08/316,424
; FILING DATE: 30-SEP-1994
; APPLICATION NUMBER: PCT/US94/03349
; FILING DATE: 29-MAR-1994
; APPLICATION NUMBER: US 08/040,942
; FILING DATE: 31-MAR-1993
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-10-581-26
Query Match 17.6%; Score 140; DB 14; Length 141;
Best Local Similarity 27.6%; Pred. No. 2.1e-07;
Matches 42; Conservative 20; Mismatches 74; Indels 16; Gaps 4;
QY 1 VLSGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKPDFR--KHLKTEAEKAS 58
Db 1 VLSAADKANVKAAGKVGQAGAHGAELERMFGLGPTTKTYFPHNLSH-----GS 52
QY 59 EDLKKHGVTVTALGAILKKKGHHE---AELKPLAQSHATKHKIPKYLEFISEAIHVL 115
Db 53 DOVKAGQKVADAL---TKAVGHLDLPGALSALSDLHAHLKLRVDPVNFKLLSHCLLVTL 109
QY 116 HSRHPDGDGADAQAGAMNKALELFRKDIAAKYK 147
Db 110 AAHPDDPNPSVHASLDKFLANVSTVLTSKYR 141
RESULT 7
US-10-289-009-4
; Sequence 4, Application US/10289009
; Publication No. US20030228700A1
; GENERAL INFORMATION:
; APPLICANT: Peters, Eric C.
; APPLICANT: Brock, Ansgar
; APPLICANT: Ericson, Christer
; APPLICANT: IRM LLC
; TITLE OF INVENTION: Labeling Reagent and Methods of Use
; FILE REFERENCE: 021288-00023005
; WOLPE, STEPHEN D.
; TITLE OF INVENTION: HEMOGLOBIN ALPHA CHAIN PEPTIDE FRAGMENTS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 MB disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/463,699
; FILING DATE: 18-Jun-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 10/128,581
; FILING DATE: 24-APR-2002
; APPLICATION NUMBER: US 09/617,840
; FILING DATE: 17-JUL-2000
; APPLICATION NUMBER: US 09/005,546
; FILING DATE: 12-JAN-1998
; APPLICATION NUMBER: US 08/535,882
; FILING DATE: 28-SEP-1995
; APPLICATION NUMBER: US 08/316,424
; FILING DATE: 30-SEP-1994
; APPLICATION NUMBER: PCT/US94/03349
; FILING DATE: 29-MAR-1994
; APPLICATION NUMBER: US 08/040,942
; FILING DATE: 31-MAR-1993
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-10-463-699-28
Query Match 17.6%; Score 140; DB 16; Length 141;
Best Local Similarity 27.6%; Pred. No. 2.1e-07;
Matches 42; Conservative 20; Mismatches 74; Indels 16; Gaps 4;
QY 1 VLSGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKPDFR--KHLKTEAEKAS 58
Db 1 VLSAADKANVKAAGKVGQAGAHGAELERMFGLGPTTKTYFPHNLSH-----GS 52
QY 59 EDLKKHGVTVTALGAILKKKGHHE---AELKPLAQSHATKHKIPKYLEFISEAIHVL 115
Db 53 DOVKAGQKVADAL---TKAVGHLDLPGALSALSDLHAHLKLRVDPVNFKLLSHCLLVTL 109
QY 116 HSRHPDGDGADAQAGAMNKALELFRKDIAAKYK 147
Db 110 AAHPDDPNPSVHASLDKFLANVSTVLTSKYR 141
```

```
; CURRENT APPLICATION NUMBER: US/10/289,009
; CURRENT FILING DATE: 2003-04-01
; PRIOR APPLICATION NUMBER: US 60/332,988
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 60/385,835
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 60/410,382
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: equine
; OTHER INFORMATION: myoglobin tryptic polypeptide #3
US-10-289-009-4

Query Match          16.8%; Score 133; DB 15; Length 31;
Best Local Similarity 76.7%; Pred. No. 1.6e-07;
Matches 23; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 LSGGEWQLVHVWAKVEADVAGHGQDILIR 31
Db 2 LSDGEWQVLNVWKGVEADVAGHGQEVILR 31

RESULT 8
US-10-378-029-102
; Sequence 102, Application US/10378029
; Publication No. US20040014087A1
; GENERAL INFORMATION:
; APPLICANT: HODGSON, David M.; LINCOLN, Stephen E.
; APPLICANT: RUSSO, Frank D.; SPIRO, Peter A.
; APPLICANT: BANVILLE, Steve C.; BRATCHER, Shawn R.
; APPLICANT: DUFOUR, Gerard E.; COHEN, Howard J.
; APPLICANT: ROSEN, Bruce; CHALUP, Michael S.
; APPLICANT: JACKSON, Jennifer L.; JONES, Annisa L.
; APPLICANT: YU, Jimmy Y.; GREENAWALT, Lila B.
; APPLICANT: PANZER, Scott R.; ROSEBERRY LINCOLN, Ann M.
; APPLICANT: WRIGHT, Rachel J.; DANIELS, Susan E.
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PT-1022-1 CIP
; CURRENT APPLICATION NUMBER: US/10/378,029
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 09/980,285
; PRIOR FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: PCT/US00/15404
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: US 60/147,500
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: US 60/147,542
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: US 60/147,541
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: US 60/147,824
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: US 60/147,547
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: US 60/147,530
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: US 60/147,536
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: US 60/147,520
; PRIOR FILING DATE: 1999-08-05
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 104
; SOFTWARE: PERL Program
; SEQ ID NO 102
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040014087A1 441779.1.j.orf1
US-10-378-029-102

Query Match          16.8%; Score 133; DB 15; Length 154;
Best Local Similarity 26.7%; Pred. No. 1.4e-06;
Matches 40; Conservative 24; Mismatches 74; Indels 12; Gaps 3;

QY 1 VLSEGEWQLVHVWAKVEADVAGHGQDILIRLPSHPETLEKFDREKHLKTEAEWKASED 60
Db 14 VLSAADKTNIKNCWKIGGEGYGEALQRMFAFPPTTKTYFSDHIDVSPGSAQVKA--- 70
QY 61 LKKHGVTVLTALGAILKKKGHE---AELKPLAQSHATKHKIPIKYLEFISEAIIHVLS 117
Db 71 ---HGKKVADALA---KAADHVEDLPGALSTLSDIHAHKLKRVDPVNFKLSHCLLVTLAC 124
QY 118 RHPGDFGADAQGMNKALELFRKIDIAAKYK 147
Db 125 HHPGDFTPAMHASLDKFLASVSTVLTSKYR 154

RESULT 9
US-10-424-599-227247
; Sequence 227247, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 227247
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_47233C.1.pep
US-10-424-599-227247

Query Match          16.6%; Score 132; DB 12; Length 142;
Best Local Similarity 26.7%; Pred. No. 1.7e-06;
Matches 40; Conservative 24; Mismatches 74; Indels 12; Gaps 3;

QY 1 VLSEGEWQLVHVWAKVEADVAGHGQDILIRLPSHPETLEKFDREKHLKTEAEWKASED 60
Db 2 VLSADDKTNIKNCWKIGGEGYGEALQRMFAFPPTTKTYFSDHIDVSPGSAQVKA--- 58
QY 61 LKKHGVTVLTALGAILKKKGHE---AELKPLAQSHATKHKIPIKYLEFISEAIIHVLS 117
Db 59 ---HGKKVADALA---KAADHVEDLPGALSTLSDIHAHKLKRVDPVNFKLSHCLLVTLAC 112
QY 118 RHPGDFGADAQGMNKALELFRKIDIAAKYK 147
Db 113 HHPGDFTPAMHASLDKFLASVSTVLTSKYR 142

RESULT 10
US-10-205-219-189
; Sequence 189, Application US/10205219
; Publication No. US20030138803A1
; GENERAL INFORMATION:
; APPLICANT: Warner-Lambert Company
; APPLICANT: Lee, Kevin
; APPLICANT: Dixon, Alistair
; APPLICANT: Brooksbank, Robert
; APPLICANT: Pinnock, Robert
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
```

```
; FILE REFERENCE: WL-A-018200
; CURRENT APPLICATION NUMBER: US/10/205,219
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: GB 0118354.0
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 189
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Globin, alpha, major
US-10-205-219-189

Query Match          16.6%; Score 132; DB 14; Length 142;
Best Local Similarity 26.7%; Pred. No. 1.7e-06;
Matches 40; Conservative 24; Mismatches 74; Indels 12; Gaps 3;

QY 1 VLSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFRF--KHLKTEAEMKASED 60
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 VLSADDTNKNCKWGKIGGCGGEGEALQRMFAAFTTKTYFESHIDVSPGSAQVKA--- 58

QY 61 LKHGVTULTALGAILKKKGHE---AELKPLAQSHATKHKIPKYLEFISPAIIHVLHS 117
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 ---HGKQVADALA---KAADHVEDLPGLSTLSDLHAHKLKRVDPVNFKFLSHCLLVTLAC 112

QY 118 RHPGDFGADAGAMNKALELFRKDAIAKYK 147
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 113 HHPGDTTAMHASLDKFLASVSTVLTISKYR 142

RESULT 11
US-09-839-164-5
; Sequence 5, Application US/09839164
; Patent No. US20020098583A1
; GENERAL INFORMATION:
; APPLICANT: KOZLOV, VLADIMIR
; TSYRLOVA, IRENA
; TITLE OF INVENTION: INHIBITOR OF STEM CELL PROLIFERATION AND
; USES THEREOF
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8th FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 Mb diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 09/839,164
; FILING DATE: 23-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,668
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/316,424
; FILING DATE: 30-SEP-1994
; APPLICATION NUMBER: PCT/US94/03349
; FILING DATE: 29-MAR-1994
; APPLICATION NUMBER: US 08/040,942
; FILING DATE: 31-MAR-1993
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-10-128-581-24
; Sequence 24, Application US/10128581
; Publication No. US20030104984A1
; GENERAL INFORMATION:
; APPLICANT: TSYRLOVA, IRENA
; WOLPE, STEPHEN D.
; TITLE OF INVENTION: HEMOGLOBIN ALPHA CHAIN PEPTIDE FRAGMENTS USEFUL
; FOR INHIBITING STEM CELL PROLIFERATION
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/128,581
; FILING DATE: 24-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/617,840
; FILING DATE: 17-JUL-2000
; APPLICATION NUMBER: US 09/005,546
; FILING DATE: 12-JAN-1998
; APPLICATION NUMBER: US 08/535,882
; FILING DATE: 28-SEP-1995
; APPLICATION NUMBER: US 08/316,424
; FILING DATE: 30-SEP-1994
; APPLICATION NUMBER: PCT/US94/03349
; FILING DATE: 29-MAR-1994
; APPLICATION NUMBER: US 08/040,942
; FILING DATE: 31-MAR-1993
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 24:

US-10-128-581-24
; Query Match          16.1%; Score 128; DB 14; Length 141;
; Best Local Similarity 27.0%; Pred. No. 4.6e-06;
; Matches 41; Conservative 22; Mismatches 73; Indels 16; Gaps 4;

QY 1 VLSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFRF--KHLKTEAEMKAS 58
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 VLSGEDKSNIAKAWGKIGGHGAEGAEALERMFASPTTKTYFPHFDVSH-----GS 52

QY 59 EDLKKHGVTLTALGAILKKKGHE---AELKPLAQSHATKHKIPKYLEFISEAIIHVL 115
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 53 AQVKGHGKQVADALAS---AAGHLDDLPGLSALSLSDLHAHKLKRVDPVNFKLLSHCLLVTL 109

QY 116 HSRHFGDAGADAGAMNKALELFRKDAIAKYK 147
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 110 ASHHPADFTPAVHASLDKFLASVSTVLTISKYR 141

RESULT 12
US-10-128-581-24
; Sequence 24, Application US/10128581
; Publication No. US20030104984A1
; GENERAL INFORMATION:
; APPLICANT: TSYRLOVA, IRENA
; WOLPE, STEPHEN D.
; TITLE OF INVENTION: HEMOGLOBIN ALPHA CHAIN PEPTIDE FRAGMENTS USEFUL
; FOR INHIBITING STEM CELL PROLIFERATION
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/128,581
; FILING DATE: 24-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/617,840
; FILING DATE: 17-JUL-2000
; APPLICATION NUMBER: US 09/005,546
; FILING DATE: 12-JAN-1998
; APPLICATION NUMBER: US 08/535,882
; FILING DATE: 28-SEP-1995
; APPLICATION NUMBER: US 08/316,424
; FILING DATE: 30-SEP-1994
; APPLICATION NUMBER: PCT/US94/03349
; FILING DATE: 29-MAR-1994
; APPLICATION NUMBER: US 08/040,942
; FILING DATE: 31-MAR-1993
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 24:

US-10-128-581-24
; Query Match          16.1%; Score 128; DB 14; Length 141;
; Best Local Similarity 27.0%; Pred. No. 4.6e-06;
; Matches 41; Conservative 22; Mismatches 73; Indels 16; Gaps 4;
```

```

Qy 1 VLSGEWOLVUHVWAKVADVAGHQQDILIRLQKSHPETLEKPDRF--KHLKTEAMKAS 58
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 VLSGEDKSNIAAWGKIGHGAEYGAELERMFASFPTTKYFPFVDSH-----GS 52
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 59 EDLKKHGVTVLTALGAILKKGHHE---AELPLAQSHATKUKIPKYLEFISEALIHVL 115
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 53 AQVKGHGKKVADALAS---AAGHLDDLPGALSJDLHAHKLRDPVNFPLSHCILLVL 109
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 116 HSRHPGDDGADAQAGAMNKALELFRKDIAAKYK 147
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 110 ASHPADPTPAVHASIDFLASVSTVLTISKYR 141
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 13
 US-10-463-699-26
 ; Sequence 26, Application US/10463699
 ; Publication No. US20040081640A1
 ; GENERAL INFORMATION:
 ; APPLICANT: TSVRLOVA, IRENA
 ; ; WOLPE, STEPHEN D.
 ; TITLE OF INVENTION: HEMOGLOBIN ALPHA CHAIN PEPTIDE FRAGMENTS
 ; ; USEFUL FOR INHIBITING STEM CELL PROLIFERATION
 ;
 ; NUMBER OF SEQUENCES: 33
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: NIXON & VANDERHYE P.C.
 ; STREET: 1100 NORTH GLEBE ROAD
 ; CITY: ARLINGTON
 ; STATE: VIRGINIA
 ; COUNTRY: U.S.A.
 ; ZIP: 22201-4714
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 1.44 MB Disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: MS word
 ;
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/463,699
 ; FILING DATE: 18-Jun-2003
 ; CLASSIFICATION: <Unknown>
 ;
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 10/128,581
 ; FILING DATE: 24-APR-2002
 ; APPLICATION NUMBER: US 09/617,840
 ; FILING DATE: 17-JUL-2000
 ; APPLICATION NUMBER: US 09/005,546
 ; FILING DATE: 12-JAN-1998
 ; APPLICATION NUMBER: US 08/535,882
 ; FILING DATE: 28-SEP-1995
 ; APPLICATION NUMBER: US 08/316,424
 ; FILING DATE: 30-SEP-1994
 ; APPLICATION NUMBER: PCT/US94/03349
 ; FILING DATE: 29-MAR-1994
 ; APPLICATION NUMBER: US 08/040,942
 ; FILING DATE: 31-MAR-1993
 ;
 ; INFORMATION FOR SEQ ID NO: 26:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 141 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: <Unknown>
 ; TOPOLOGY: linear
 ;
 ; MOLECULE TYPE: peptide
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 26:
 ;
 ; US-10-463-699-26

```

Query_Match      16.1%; Score 128; DB 16; Length 141;
Best Local Similarity 27.0%; Pred. No. 4.6e-06;
Matches 41; Conservative 22; Mismatches 73; Indels 16; Gaps 4;

Qy 1 VLSEGEWLVLVHAKVEADVAGHGQDILILFKSHPTLEKFORP--KHLKTAEMKAS 58
Db 1 VLSSGDKNIIKAAGWKIGGHGAEYGAELERMFASFPTTKTYFHFDPVSH-----GS 52

```

Qy	59	EDLKKHGVTVTALGAILKKKGHHH----	AELKPLAQSHATKHIPILKYLEFSEALIHVL	115
Db	53	AQVKGHGKKVADALAS----	AAGLHDDLPGALASJDLHAHKLVDPVNFKLSSHCLLVTL	109
Qy	116	HSRHPGDFGADAQAGAMNKALFLPRKDIAAKYK	147	
Db	110	ASHEPADFTPAVHASLQDKFLASVSVTLTSKYR	141	

RESULT 14
US-09-839-164-8
; Sequence 8, Application US/09839164
; Patent No. US20020098583A1
; GENERAL INFORMATION:
APPLICANT: KOZLOV, VLADIMIR
TSYRLOVA, IRENA
TITLE OF INVENTION: INHIBITOR OF STEM CELL PROLIFERATION AND
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEEHE ROAD, 8th FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: 1.44 Mb diskette
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MS Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/839,164
FILING DATE: 23-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/477,668
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US 08/316,424
FILING DATE: 30-SEP-1994
APPLICATION NUMBER: PCT/US94/03349
FILING DATE: 29-MAR-1994
APPLICATION NUMBER: US 08/040,942
FILING DATE: 31-MAR-1993
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 146 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-839-164-8

Query Match	16.0%;	Score 127;	DB 9;	Length 146;
Best Local Similarity	27.6%;	Pred. No. 6.2e-06;		
Matches	40;	Conservative 20;	Mismatches 83;	Indels 2;
Gaps	1			
QY	2	LSGEWQLVLHWAKVEADVAGHGODILIRLFKSHPETLEKDPDRKHLKTEARMKASEDL	61	
DB	3	LSAEKEAVLGGWKVNVDEV--GGELGRLLVVPTQTFPESFGDLSNADAVMGPKV	60	
QY	62	KKHGVTVLTALGAILKKKHGHEAEELKPLAQSHATKHKIPKYLEFISEAIIHVLHSHRHPG	121	
DB	61	KAHGKVKVLSQSFSDGLKHLNDLNGTKTAKLSELHCDQLHVDPENFRLLGNIVVVVLRRLGH	120	
QY	122	DFGADAQGAMNKALELFRKDIAAKY	146	
DB	121	DFNPDPQAAFOKVVAGVANALAHKY	145	

RESULT 15
US-10-128-581-27

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:26:21 ; Search time 94.2821 Seconds
(without alignments)
1583.930 Million cell updates/sec

Title: US-09-455-978B-76
Perfect score: 794
Sequence: 1 VLSEGEWLVLHVAKVEAD.....ALELFRKDIAKYKELGYQG 153

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA Main:*

- 1: /cgn2_6/ptodata/2/paa/PTCUS_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/2/paa/US097A_COMB.pep.*
- 22: /cgn2_6/ptodata/2/paa/US097B_COMB.pep.*
- 23: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
- 24: /cgn2_6/ptodata/2/paa/US099A_COMB.pep.*
- 25: /cgn2_6/ptodata/2/paa/US099B_COMB.pep.*
- 26: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
- 27: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
- 28: /cgn2_6/ptodata/2/paa/US102_COMB.pep.*
- 29: /cgn2_6/ptodata/2/paa/US103_COMB.pep.*
- 30: /cgn2_6/ptodata/2/paa/US104_COMB.pep.*
- 31: /cgn2_6/ptodata/2/paa/US106_COMB.pep.*
- 32: /cgn2_6/ptodata/2/paa/US107_COMB.pep.*
- 33: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	ID	Description

1	794	100.0	153	18	US-09-455-978B-76	Sequence 76, Appli
2	794	100.0	153	21	US-09-700-708-4	Sequence 4, Appli
3	794	100.0	154	22	US-09-791-537-51499	Sequence 51499, A
4	789	99.4	153	22	US-09-791-537-41444	Sequence 41444, A
5	789	99.4	154	22	US-09-791-537-23678	Sequence 23678, A
6	788	99.2	154	22	US-09-791-537-13857	Sequence 13857, A
7	788	99.2	154	22	US-09-791-537-22569	Sequence 22569, A
8	786	99.0	154	22	US-09-791-537-23088	Sequence 23088, A
9	786	99.0	154	22	US-09-791-537-23776	Sequence 23776, A
10	786	99.0	154	22	US-09-791-537-102872	Sequence 102872, A
11	785	98.9	154	22	US-09-791-537-22535	Sequence 22535, A
12	785	98.9	154	22	US-09-791-537-23775	Sequence 23775, A
13	785	98.9	154	22	US-09-791-537-102151	Sequence 102151, A
14	785	98.9	154	22	US-09-791-537-102898	Sequence 102898, A
15	784	98.7	153	22	US-09-791-537-51498	Sequence 51498, A
16	784	98.7	154	22	US-09-791-537-13855	Sequence 13855, A
17	784	98.7	154	22	US-09-791-537-17232	Sequence 17232, A
18	784	98.7	154	22	US-09-791-537-22555	Sequence 22555, A
19	784	98.7	154	22	US-09-791-537-102870	Sequence 102870, A
20	783	98.6	151	22	US-09-791-537-71491	Sequence 71491, A
21	783	98.6	154	22	US-09-791-537-20784	Sequence 20784, A
22	783	98.6	154	22	US-09-791-537-33391	Sequence 33391, A
23	783	98.6	154	22	US-09-791-537-74339	Sequence 74339, A
24	783	98.6	154	22	US-09-791-537-102153	Sequence 102153, A
25	783	98.6	154	22	US-09-791-537-102916	Sequence 102916, A
26	783	98.6	154	22	US-09-791-537-151603	Sequence 151603, A
27	782	98.5	154	22	US-09-791-537-20811	Sequence 20811, A
28	782	98.5	154	22	US-09-791-537-73835	Sequence 73835, A
29	782	98.5	154	22	US-09-791-537-102896	Sequence 102896, A
30	781	98.4	154	22	US-09-791-537-13847	Sequence 13847, A
31	781	98.4	154	22	US-09-791-537-23671	Sequence 23671, A
32	781	98.4	154	22	US-09-791-537-102150	Sequence 102150, A
33	780	98.2	154	22	US-09-791-537-67801	Sequence 67801, A
34	780	98.2	154	22	US-09-791-537-102163	Sequence 102163, A
35	779	98.1	154	22	US-09-791-537-23158	Sequence 23158, A
36	779	98.1	154	22	US-09-791-537-23675	Sequence 23675, A
37	779	98.1	154	22	US-09-791-537-27099	Sequence 27099, A
38	778	98.0	151	22	US-09-791-537-108163	Sequence 108163, A
39	778	98.0	154	22	US-09-791-537-23159	Sequence 23159, A
40	778	98.0	154	22	US-09-791-537-23676	Sequence 23676, A
41	778	98.0	154	22	US-09-791-537-102159	Sequence 102159, A
42	775	97.6	154	22	US-09-791-537-23089	Sequence 23089, A
43	775	97.4	153	22	US-09-791-537-9291	Sequence 9291, Ap
44	772	97.2	154	22	US-09-791-537-27073	Sequence 27073, A
45	772	97.2	154	22	US-09-791-537-77660	Sequence 77660, A

ALIGNMENTS

RESULT 1
US-09-455-978B-76
; Sequence 76, Application US/09455978B
; GENERAL INFORMATION:
; APPLICANT: Alam, Magsudul
; APPLICANT: Larsen, Randy
; TITLE OF INVENTION: HEME PROTEINS HEMAT-HS AND HEMAT-BS AND THEIR USE IN
; FILE REFERENCE: 201040/1020
; CURRENT APPLICATION NUMBER: US/09/455,978B
; CURRENT FILING DATE: 1999-12-06
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 76
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Sperm-whale myoglobin
US-09-455-978B-76

Query Match 100.0%; Score 794; DB 18; Length 153;
Best Local Similarity 100.0%; Pred. No. 2.4e-77;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRPHLKTAEAMKASED	60
Db	1	VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRPHLKTAEAMKASED	60
Qy	61	LKKGHTVTLTALGAILKKKGHHHEAEKPLAQSHATKHKIPIKYLFPISEAILHVLSHRHP	120
Db	61	LKKGHTVTLTALGAILKKKGHHHEAEKPLAQSHATKHKIPIKYLFPISEAILHVLSHRHP	120
Qy	121	GDFGADAQAGAMNKALELFRKDIAAKYKELGYOG	153
Db	121	GDFGADAQAGAMNKALELFRKDIAAKYKELGYOG	153

RESULT 2

```

US-09-700-708-4
; Sequence 4, Application US/09700708
; GENERAL INFORMATION:
; APPLICANT: ITAI, Akiko
; APPLICANT: ITAI, Reiko
; APPLICANT: TOMIOKA, Nobuo
; TITLE OF INVENTION: Method For Predicting Functions of Protein
; FILE REFERENCE: P20234
; CURRENT APPLICATION NUMBER: US/09/700,708
; CURRENT FILING DATE: 2000-11-24
; PRIOR APPLICATION NUMBER: PCT/JP98/02302
; PRIOR FILING DATE: 1998-05-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Whale
US-09-700-708-4

```

RESULT 3

```

US-09-791-537-51499
; Sequence 51499, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAM
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 51499
; LENGTH: 154
; TYPE: prt
; ORGANISM: pdb 1DUKA
US-09-791-537-51499

```

	Matches	153;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	VLSEGEQQLVLHWAKV	ADVAGHGQDILRLFKSHPETLEK	DFRDKHLKTEAE	KASD	60				
Db	2	VLSEGEQQLVLHWAKV	ADVAGHGQDILRLFKSHPETLEK	DFRDKHLKTEAE	KASD	61				
Qy	61	LKKHGVTVLTGAILKKKG	HHEAELKPLAQSHATKHKIP	KYLEPTSEAI	IHVLSHRHP	120				
Db	62	LKKHGVTVLTGAILKKKG	HHEAELKPLAQSHATKHKIP	KYLEPTSEAI	IHVLSHRHP	121				
Qy	121	GDFGADAQGAMNKALEL	FRKDIAAKYKEIGYQG	153						
Db	122	GDFGADAQGAMNKALEL	FRKDIAAKYKEIGYQG	154						

RESULT 4

```

RES001 4
US-09-791-537-41444
; Sequence 41444, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Biomomix, Inc.
; APPLICANT: Danzer, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURE
; FILE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 41444
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Physter catodon
US-09-791-537-41444

```

RESULT. T 5

```

US-09-791-537-23678
; Sequence 23678, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Bionomix, Inc.
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURAL
; FILE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 23678
; LENGTH: 154
; TYPE: PRT
; ORGANISM: pdb 2MGK
US-09-791-537-23678

```

Best Local Similarity 99.3%; Pred. No. 8.5e-77;
Matches 152; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASED 60
Db 2 VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASED 61
QY 61 LKKHGVTVLTALGAILKKKGHHEAEKPLAQSHATKHKPIKYLEFISEAIIHVLHSRHP 120
Db 62 LKKHGVTVLTALGAILKKKGHHEAEKPLAQSHATKHKPIKYLEFISEAIIHVLHSRHP 121
QY 121 GDFGADAQGMNKALELFRKDIKAAKYKELGYQG 153
Db 122 GNFADAQGMNKALELFRKDIKAAKYKELGYQG 154

RESULT 6

US-09-791-537-13857
; Sequence 13857, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13857
; LENGTH: 154
; TYPE: PRT
; ORGANISM: pdb ICIOA
US-09-791-537-13857

Query Match 99.2%; Score 788; DB 22; Length 154;
Best Local Similarity 98.7%; Pred. No. 1.1e-76;
Matches 151; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASED 60
Db 2 VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASED 61
QY 61 LKKHGVTVLTALGAILKKKGHHEAEKPLAQSHATKHKPIKYLEFISEAIIHVLHSRHP 120
Db 62 LKKHGVTVLTALGAILKKKGHHEAEKPLAQSHATKHKPIKYLEFISEAIIHVLHSRHP 121
QY 121 GDFGADAQGMNKALELFRKDIKAAKYKELGYQG 153
Db 122 GNFADAQGMNKALELFRKDIKAAKYKELGYQG 154

RESULT 7

US-09-791-537-22569
; Sequence 22569, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22569
; LENGTH: 154
; TYPE: PRT
; ORGANISM: pdb IMLM
US-09-791-537-22569

Query Match 99.2%; Score 788; DB 22; Length 154;
Best Local Similarity 98.7%; Pred. No. 1.1e-76;
Matches 151; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASED 60
Db 2 VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASED 61
QY 61 LKKHGVTVLTALGAILKKKGHHEAEKPLAQSHATKHKPIKYLEFISEAIIHVLHSRHP 120
Db 62 LKKHGVTVLTALGAILKKKGHHEAEKPLAQSHATKHKPIKYLEFISEAIIHVLHSRHP 121
QY 121 GDFGADAQGMNKALELFRKDIKAAKYKELGYQG 153
Db 122 GNFADAQGMNKALELFRKDIKAAKYKELGYQG 154

RESULT 8

US-09-791-537-23088
; Sequence 23088, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23088
; LENGTH: 154
; TYPE: PRT
; ORGANISM: pdb IMLR
US-09-791-537-23088

Query Match 99.0%; Score 786; DB 22; Length 154;
Best Local Similarity 98.7%; Pred. No. 1.8e-76;
Matches 151; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASED 60
Db 2 VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEWKASED 61
QY 61 LKKHGVTVLTALGAILKKKGHHEAEKPLAQSHATKHKPIKYLEFISEAIIHVLHSRHP 120
Db 62 LKKHGVTVLTALGAILKKKGHHEAEKPLAQSHATKHKPIKYLEFISEAIIHVLHSRHP 121
QY 121 GDFGADAQGMNKALELFRKDIKAAKYKELGYQG 153
Db 122 GNFADAQGMNKALELFRKDIKAAKYKELGYQG 154

RESULT 9

US-09-791-537-23776
; Sequence 23776, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23776
; LENGTH: 154
; TYPE: PRT
; ORGANISM: pdb 2SPO
US-09-791-537-23776


```
; ORGANISM: pdb 1CH2A
US-09-791-537-102151

Query Match      98.9%; Score 785; DB 22; Length 154;
Best Local Similarity 98.7%; Pred. No. 2.3e-76;
Matches 151; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKFDKPKHLKTEAEKMSD 60
DB 2 VLSEGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKFDKPKHLKTEAEKMSD 61
QY 61 LKKHGVTVLTALGAILKKKGHEAEELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 120
DB 62 LKKHGVTVLTALGAILKKKGHEAEELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 121
QY 121 GDFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
DB 122 GNFGADAQAGAMNKALELFRKDIAAKYKELGYQG 154

RESULT 14
US-09-791-537-102898
; Sequence 102898, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; FILE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 102898
; LENGTH: 154
; TYPE: PRT
; ORGANISM: pdb 1CP5A
US-09-791-537-102898

Query Match      98.9%; Score 785; DB 22; Length 154;
Best Local Similarity 98.7%; Pred. No. 2.3e-76;
Matches 151; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKFDKPKHLKTEAEKMSD 60
DB 2 VLSEGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKFDKPKHLKTEAEKMSD 61
QY 61 LKKHGVTVLTALGAILKKKGHEAEELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 120
DB 62 LKKHGVTVLTALGAILKKKGHEAEELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 121
QY 121 GDFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
DB 122 GNFGADAQAGAMNKALELFRKDIAAKYKELGYQG 154

RESULT 15
US-09-791-537-51498
; Sequence 51498, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; FILE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 51498
; LENGTH: 153
```

```
; TYPE: PRT
; ORGANISM: pdb 1DTWA
US-09-791-537-51498

Query Match      98.7%; Score 784; DB 22; Length 153;
Best Local Similarity 99.3%; Pred. No. 3e-76;
Matches 152; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKFDKPKHLKTEAEKMSD 60
DB 1 VLSEGEWQLVHLVWAKVEADVAGHGQDILIRLFKSHPETLEKFDKPKHLKTEAEKMSD 60
QY 61 LKKHGVTVLTALGAILKKKGHEAEELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 120
DB 61 LKKHGVTVLTALGAILKKKGHEAEELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP 120
QY 121 GDFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
DB 121 GDFGADAQAGAMNKALELFRKDIAAKYKELGYQG 153
```

Search completed: August 10, 2004, 15:39:00
Job time : 95.2821 secs

B/ANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:26:56 ; Search time 12.04 Seconds
(without alignments)
1368.899 Million cell updates/sec

Title: US-09-455-978b-76
Perfect score: 794
Sequence: 1 VLSEGEQLVLHVAKVEAD.....ALELFRKDIAAKYKELGYQG 153

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 618821

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA New:*
1: /cgn2_6/ptodata/2/paa/PCT NEW COMB.pcp.*
2: /cgn2_6/ptodata/2/paa/US06 NEW COMB.pcp.*
3: /cgn2_6/ptodata/2/paa/US07 NEW COMB.pcp.*
4: /cgn2_6/ptodata/2/paa/US08 NEW COMB.pcp.*
5: /cgn2_6/ptodata/2/paa/US09 NEW COMB.pcp.*
6: /cgn2_6/ptodata/2/paa/US10 NEW COMB.pcp.*
7: /cgn2_6/ptodata/2/paa/US60 NEW COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	699	88.0	153	6	US-10-784-004-632		Sequence 632, App
2	699	88.0	153	6	US-10-784-004-1039		Sequence 1039, App
3	699	88.0	154	6	US-10-170-205E-32946		Sequence 32946, A
4	699	88.0	154	6	US-10-170-205E-33330		Sequence 33330, A
5	699	88.0	154	6	US-10-170-205E-33331		Sequence 33331, A
6	647	81.5	154	6	US-10-784-004-310		Sequence 310, App
7	647	81.5	154	6	US-10-784-004-898		Sequence 898, App
8	472	59.4	113	6	US-10-793-479-4103		Sequence 4103, App
9	327	41.2	79	6	US-10-425-115-279565		Sequence 279565
10	314	39.5	71	6	US-10-793-479-4102		Sequence 4102, App
11	212.5	26.8	190	1	PCT-US04-07412-750		Sequence 750, App
12	212.5	26.8	190	1	PCT-US02-17443-214		Sequence 214, App
13	212.5	26.8	190	1	PCT-US02-17443-216		Sequence 216, App
14	212.5	26.8	190	1	US-10-389-559-750		Sequence 750, App
15	212.5	26.8	190	6	US-10-170-205E-19233		Sequence 19233, A
16	159	20.0	142	6	US-10-170-205E-7507		Sequence 7507, App
17	140	17.6	141	6	US-10-776-172-21		Sequence 21, App
18	140	17.6	141	6	US-10-897-005-28		Sequence 28, App
19	128	16.1	141	6	US-10-776-172-19		Sequence 19, App
20	128	16.1	141	6	US-10-897-005-26		Sequence 26, App
21	127	16.0	146	6	US-10-776-172-22		Sequence 22, App
22	127	16.0	146	6	US-10-897-005-29		Sequence 29, App
23	125	15.7	147	6	US-10-170-205E-23524		Sequence 23524, A
24	125	15.7	147	7	US-60-563-440-970		Sequence 970, App
25	124	15.6	120	6	US-10-425-115-205641		Sequence 205641, App
26	124	15.6	147	1	PCT-US03-36002-201		Sequence 201, App

ALIGNMENTS

RESULT 1

US-10-784-004-632
; Sequence 632, Application US/10784004
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784,004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 632
; LENGTH: 153
; TYPE: PRT
; ORGANISM: human
US-10-784-004-632

Query Match 88.0%; Score 699; DB 6; Length 153;
Best Local Similarity 84.9%; Pred No. 4.7e-64;
Matches 129; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY	2	LSEGEQLVLHVAKVEADVAGHQDILIRLFKSHPTLEKFDREKHLKTEAMKASEDL	61
Db	2	LSDGEQLVLNVWGKVEADIPGHQGVILIRLFKGHPTLEKFDREKHLKSEDEMKASEDL	61
QY	62	KHGVTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPG	121
Db	62	KHGVTVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPKYLEFISECIIQVLSQKHPG	121
QY	122	DFGADAQAMKALELFRKDIAAKYKELGYQG	153
Db	122	DFGADAQAMKALELFRKDMASNYKELGPFQG	153

RESULT 2

US-10-784-004-1039
; Sequence 1039, Application US/10784004
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784,004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1039
; LENGTH: 153
; TYPE: PRT
; ORGANISM: human

US-10-784-004-1039

Query Match 88.0%; Score 699; DB 6; Length 153;
Best Local Similarity 84.9%; Pred. No. 4.7e-64;
Matches 129; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY 2 LSEGEQVLVHWAKVEADVAGHGQDILIRLPKSHPETLEKDFRPHLKTAEAMKASEDL 61
DB 2 LSDGEQVLVNWGVKVEADIPGHQEVILIRLPKSHPETLEKDFRPHLKTAEAMKASEDL 61
QY 62 KKHGVTVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFTSEAIHVLHSRHPG 121
DB 62 KKHGATVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFTSEAIHVLHSRHPG 121

QY 122 DFGADAQAGMNAKLELFRKDIKAAKYKELGYQG 153
DB 122 DFGADAQAGMNAKLELFRKDMASNYKELGFQG 153

RESULT 3

US-10-170-205E-32946
; Sequence 32946, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 32946
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-32946

Query Match 88.0%; Score 699; DB 6; Length 154;
Best Local Similarity 84.9%; Pred. No. 4.7e-64;
Matches 129; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY 2 LSEGEQVLVHWAKVEADVAGHGQDILIRLPKSHPETLEKDFRPHLKTAEAMKASEDL 61
DB 3 LSDGEQVLVNWGVKVEADIPGHQEVILIRLPKSHPETLEKDFRPHLKTAEAMKASEDL 62
QY 62 KKHGVTVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFTSEAIHVLHSRHPG 121
DB 63 KKHGATVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFTSEAIHVLHSRHPG 122

QY 122 DFGADAQAGMNAKLELFRKDIKAAKYKELGYQG 153
DB 123 DFGADAQAGMNAKLELFRKDMASNYKELGFQG 154

RESULT 4

US-10-170-205E-33330
; Sequence 33330, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 33330
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-33330

Query Match 88.0%; Score 699; DB 6; Length 154;

Best Local Similarity 84.9%; Pred. No. 4.7e-64;
Matches 129; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY 2 LSEGEQVLVHWAKVEADVAGHGQDILIRLPKSHPETLEKDFRPHLKTAEAMKASEDL 61
DB 3 LSDGEQVLVNWGVKVEADIPGHQEVILIRLPKSHPETLEKDFRPHLKTAEAMKASEDL 62
QY 62 KKHGVTVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFTSEAIHVLHSRHPG 121
DB 63 KKHGATVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFTSEAIHVLHSRHPG 122
QY 122 DFGADAQAGMNAKLELFRKDIKAAKYKELGYQG 153
DB 123 DFGADAQAGMNAKLELFRKDMASNYKELGFQG 154

RESULT 5

US-10-170-205E-33331
; Sequence 33331, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 33331
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-33331

Query Match 88.0%; Score 699; DB 6; Length 154;
Best Local Similarity 84.9%; Pred. No. 4.7e-64;
Matches 129; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY 2 LSEGEQVLVHWAKVEADVAGHGQDILIRLPKSHPETLEKDFRPHLKTAEAMKASEDL 61
DB 3 LSDGEQVLVNWGVKVEADIPGHQEVILIRLPKSHPETLEKDFRPHLKTAEAMKASEDL 62
QY 62 KKHGVTVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFTSEAIHVLHSRHPG 121
DB 63 KKHGATVLTALGAILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFTSEAIHVLHSRHPG 122
QY 122 DFGADAQAGMNAKLELFRKDIKAAKYKELGYQG 153
DB 123 DFGADAQAGMNAKLELFRKDMASNYKELGFQG 154

RESULT 6

US-10-784-004-310
; Sequence 310, Application US/10784004
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784,004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 310
; LENGTH: 154
; TYPE: PRT
; ORGANISM: rat
US-10-784-004-310

Query Match 81.5%; Score 647; DB 6; Length 154;
Best Local Similarity 78.9%; Pred. No. 1e-58;
Matches 120; Conservative 18; Mismatches 14; Indels 0; Gaps 0;

QY 2 LSEGEQVLVHWAKVEADVAGHGQDILIRLPKSHPETLEKDFRPHLKTAEAMKASEDL 61

Db 3 LSDGEWQVLTWGWKVEADVAGHGDILIRLFKSHPETLEKDFKPKNLKSEEMKASD 62
QY 62 KKHGVTVTALGAILKKKGHAEALKPLAQSHATKHKIPKYLEFTSEAIHVLHSRHPG 121
Db 63 KKHGCTVTALTGILTKKGQHAETIQPLAQSHATKHKIPVKYLEFTSEVIOVLKKRYSG 122
QY 122 DFGADAQAGMKALELFRKDIAAKYKELGYQG 153
Db 123 DFGADAQAGMKALELFRNDIAAKYKELGFQG 154

RESULT 7
US-10-784-004-898
; Sequence 898, Application US/10784004
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784,004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 898
; LENGTH: 154
; TYPE: PRT
; ORGANISM: rat
US-10-784-004-898

Query Match 81.5%; Score 647; DB 6; Length 154;
Best Local Similarity 78.9%; Pred. No. 1e-58;
Matches 120; Conservative 18; Mismatches 14; Indels 0; Gaps 0;
QY 2 LSEGEWQVLTWGWKVEADVAGHGDILIRLFKSHPETLEKDFKPKNLKSEEMKASD 61
Db 3 LSDGEWQVLTWGWKVEADVAGHGDILIRLFKSHPETLEKDFKPKNLKSEEMKASD 62
QY 62 KKHGVTVTALGAILKKKGHAEALKPLAQSHATKHKIPKYLEFTSEAIHVLHSRHPG 121
Db 63 KKHGCTVTALTGILTKKGQHAETIQPLAQSHATKHKIPVKYLEFTSEVIOVLKKRYSG 122

QY 122 DFGADAQAGMKALELFRKDIAAKYKELGYQG 153
Db 123 DFGADAQAGMKALELFRNDIAAKYKELGFQG 154

RESULT 8
US-10-793-479-4103
; Sequence 4103, Application US/10793479
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/10/793,479
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: US/09/513,999
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 4103
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -20..-1
; OTHER INFORMATION: score 3.5
; OTHER INFORMATION: seq WQIVLVNMGKVEA/DI
; FEATURE:

; NAME/KEY: UNSURE
; LOCATION: 84
; OTHER INFORMATION: Xaa-Phe or Ile or Leu or Met or Val
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 85
; OTHER INFORMATION: Xaa-Pro or Thr
US-10-793-479-4103
Query Match 59.4%; Score 472; DB 6; Length 113;
Best Local Similarity 85.1%; Pred. No. 6.3e-41;
Matches 86; Conservative 9; Mismatches 6; Indels 0; Gaps 0;
QY 2 LSEGEWQVLTWGWKVEADVAGHGDILIRLFKSHPETLEKDFKPKNLKSEEMKASD 61
Db 3 LSDGEWQVLTWGWKVEADVAGHGDILIRLFKSHPETLEKDFKPKNLKSEEMKASD 62
QY 62 KKHGVTVTALGAILKKKGHAEALKPLAQSHATKHKIPK 102
Db 63 KKHGATVTALTGILTKKGHAEALKPLAQSHATKHKIPVK 103

RESULT 9
US-10-425-115-279565
; Sequence 279565, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 279565
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_106543C.1.pep
US-10-425-115-279565

Query Match 41.2%; Score 327; DB 6; Length 79;
Best Local Similarity 76.6%; Pred. No. 3e-26;
Matches 59; Conservative 12; Mismatches 6; Indels 0; Gaps 0;
QY 2 LSEGEWQVLTWGWKVEADVAGHGDILIRLFKSHPETLEKDFKPKNLKSEEMKASD 61
Db 3 LSDGEWQVLTWGWKVEADVAGHGDILIRLFKSHPETLEKDFKPKNLKSEEMKASD 62
QY 62 KKHGVTVTALGAILKK 78
Db 63 KKHGCTVTALTGILTK 79

RESULT 10
US-10-793-479-4102
; Sequence 4102, Application US/10793479
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/10/793,479
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: US/09/513,999
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681

```
; SOFTWARE: Patent.pm
; SEQ ID NO 4102
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -20...-1
; OTHER INFORMATION: score 3.5
; OTHER INFORMATION: seq WQVLVNWGKVEA/DI
US-10-793-479-4102

Query Match          39.5%; Score 314; DB 6; Length 71;
Best Local Similarity 82.6%; Pred. No. 5.7e-25;
Matches 57; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEKASEDL 61
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3 LSDGEWQLVNWGKVEADPGHGQEVILIRLFKSHPETLEKDFRFXHLKTEAEKASEDL 62
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 62 KKHGVTVLT 70
   |||||:|||||
Db 63 KKHGATVLT 71
   |||||:|||||

RESULT 11
PCT-US04-07412-750
; Sequence 750, Application PC/TUS0407412
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-rui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Ghosh, Malabika
; APPLICANT: Xue, Aidong J.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Dunrui
; APPLICANT: Goodrich, Ryle W.
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wehrman, Tom
; APPLICANT: Weng, Gezhi
; APPLICANT: Wang, Zhiwei
; APPLICANT: Boyle, Bryan J.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 822CIP/PCT
; CURRENT APPLICATION NUMBER: PCT/US04/07412
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: US 10/389,559
; PRIOR FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: US 60/365,264
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US 60/340,187
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 10/296,115
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 10/275,027
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1920

; SOFTWARE: pt_FL_genes Version 6.0
; SEQ ID NO 750
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION:
PCT-US04-07412-750

Query Match          26.8%; Score 212.5; DB 1; Length 190;
Best Local Similarity 31.6%; Pred. No. 5.4e-14;
Matches 49; Conservative 31; Mismatches 68; Indels 7; Gaps 2;

QY 2 LSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEKASEDL 61
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 19 LSEAEKAVQAMWALYANCEDVGVALIVRFVNFPSAKQYFSQFKHMDPLEMERSPOL 78
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 62 KKHGVTVLTALGAILKKKHHEAE-----LKLPAQSHATKHKIPKILFEIGAIHVHL 116
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 79 RKHACRVMGALNTVVVENL--HDPDKVSSVLAIVGKAHALKHKVEPVYFKILSGVILEVVA 136
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 117 SRHPGDFGADACGAMNKALELFRKDIAAKYKELGY 151
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 137 EEFASDFPPETORAWAKLRGLIYSHVTRAYKEVGW 171
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 12
PCT-US02-17443-214
; Sequence 214, Application PC/TUS0217443
; GENERAL INFORMATION:
; APPLICANT: CuraGen Corporation
; APPLICANT: Agee, Michele
; APPLICANT: Alsobrook, John
; APPLICANT: Anderson, David
; APPLICANT: Berghs, Constance
; APPLICANT: Burgess, Catherine
; APPLICANT: Casman, Stacie
; APPLICANT: Catterton, Elina
; APPLICANT: DiPippo, Vincent
; APPLICANT: Edinger, Shlomit
; APPLICANT: Eisen, Andrew
; APPLICANT: Ellerman, Karen
; APPLICANT: Gangolli, Esha
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Herrmann, John
; APPLICANT: Hjalt, Tord
; APPLICANT: Ji, Weizhen
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Khrantsov, Nikolai
; APPLICANT: Lepley, Denise
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Malyankar, Uriel
; APPLICANT: Miller, Charles
; APPLICANT: Millet, Isabelle
; APPLICANT: Ort, Tatiana
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Pena, Carol
; APPLICANT: Rastelli, Luca
; APPLICANT: Rieger, Daniel
; APPLICANT: Rothenberg, Mark
; APPLICANT: Shenoy, Suresh
; APPLICANT: Shinkets, Richard
; APPLICANT: Smithson, Glennda
; APPLICANT: Spaderna, Steven
; APPLICANT: Spytek, Kimberly
; APPLICANT: Stone, David
; APPLICANT: Vernet, Corine
; APPLICANT: Zhong, Haihong
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-377C-061
```


US-10-389-559-750

```
; Sequence 750, Application US/10389559
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-ru
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Ghosh, Malabika
; APPLICANT: Xue, Aidong J.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Dunrui
; APPLICANT: Goodrich, Kyle W.
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Gezhi
; APPLICANT: Wang, Zhiwei
; APPLICANT: Boyle, Bryan J.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 822
; CURRENT APPLICATION NUMBER: US/10/389,559
; PRIOR FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: US 60/365,264
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US 60/340,187
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1920
; SOFTWARE: pt_FL_genes Version 6.0
; SEQ ID NO 750
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-389-559-750
```

```
Query Match 26.8%; Score 212.5; DB 6; Length 190;
Best Local Similarity 31.6%; Pred. No. 5.4e-14;
Matches 49; Conservative 31; Mismatches 68; Indels 7; Gaps 2;

QY 2 LSEGEQVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDREKHLKTEAEKASEDL 61
Db 19 LSEAEKAVQAMWARYLNCEDVGVAIIVRFFVNFPSAKQYFSQFKHMDPLEMERSPOL 78
QY 62 KKHGVTVLTALGAILKKKGHEAE-----LKPLAQSHATKHKIPKYLEFISEAIIHVLH 116
Db 79 RKHACRVMGALNTVVENL--HDPDKVSSVLALVGKAHALKHKVEPVYFKILSGVILEVVA 136
QY 117 SRHPGDFGADAQGANMKALELFRKIDIAAKYKELGY 151
Db 137 BEFASDFPPTQRAWAKLRGLIYSHVTAAYKEVGW 171
```

RESULT 15
US-10-170-205E-19233

; Sequence 19233, Application US/10170205E

```
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19233
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-19233
```

```
Query Match 26.8%; Score 212.5; DB 6; Length 190;
Best Local Similarity 31.6%; Pred. No. 5.4e-14;
Matches 49; Conservative 31; Mismatches 68; Indels 7; Gaps 2;

QY 2 LSEGEQVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDREKHLKTEAEKASEDL 61
Db 19 LSEAEKAVQAMWARYLNCEDVGVAIIVRFFVNFPSAKQYFSQFKHMDPLEMERSPOL 78
QY 62 KKHGVTVLTALGAILKKKGHEAE-----LKPLAQSHATKHKIPKYLEFISEAIIHVLH 116
Db 79 RKHACRVMGALNTVVENL--HDPDKVSSVLALVGKAHALKHKVEPVYFKILSGVILEVVA 136
QY 117 SRHPGDFGADAQGANMKALELFRKIDIAAKYKELGY 151
Db 137 BEFASDFPPTQRAWAKLRGLIYSHVTAAYKEVGW 171
```

Search completed: August 10, 2004, 15:40:16
Job time : 13.04 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:22:45 ; Search time 6.66828 Seconds
(without alignments)
2207.061 Million cell updates/sec

Title: US-09-455-978b-76

Perfect score: 794

Sequence: 1 VLSEGEWQLVHLHWAKVEAD.....ALELFRKDIAAKYKELGYQG 153

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	794	100.0	153	1 MYWHP	myoglobin [validat
2	773	97.4	153	1 MYWHM	myoglobin - dwarf
3	746	94.0	153	1 MYWHC	myoglobin - Califo
4	740	93.2	153	1 MYWHH	myoglobin - humpha
5	738	92.9	153	1 MYDDAR	myoglobin - Amazon
6	737	92.8	153	1 MYWHT	myoglobin - pilot
7	734	92.4	153	1 MYWHL	myoglobin - killer
8	733	92.3	153	1 MYDD	myoglobin - bottle
9	733	92.3	153	1 MYDDBS	myoglobin - saddle
10	731	92.1	153	1 MYPE	myoglobin - harbor
11	731	92.1	153	1 MYPED	myoglobin - Dall's
12	731	92.1	153	1 MYWHK	myoglobin - minke
13	726	91.4	153	1 MYWHE	myoglobin - finbac
14	725	91.3	153	1 MYWHU	myoglobin - Hubbs'
15	722	90.9	153	1 MYWHZ	myoglobin - goose-
16	722	90.9	153	2 JN0410	myoglobin - Europe
17	712	89.7	153	1 MYHO	myoglobin [validat
18	712	89.7	153	1 MYHOZ	myoglobin - common
19	707	89.0	153	1 MYCZ	myoglobin - chimpa
20	705	88.8	153	1 MYLP	myoglobin - potto
21	703	88.5	153	1 MYRB	myoglobin - rabbit
22	701	88.3	153	1 MYNO126	myoglobin - mounta
23	700	88.3	153	1 MYGO	myoglobin - muskra
24	700	88.2	153	1 MYGP	myoglobin [validat
25	699	88.0	154	1 MYHU	myoglobin [validat
26	695	87.5	153	1 MYQIM	myoglobin - crab-e
27	694	87.4	153	1 MYOG	myoglobin - Bornea
28	693	87.3	153	1 MYGI	myoglobin - agile
29	693	87.3	153	1 MYTS	myoglobin - common

30	693	87.3	153	2 A61366	myoglobin - siaman
31	691	87.0	153	1 MYBAO	myoglobin - olive
32	691	87.0	153	1 MYWQRG	myoglobin - red gu
33	691	87.0	153	1 MYMQHL	myoglobin - hanuma
34	691	87.0	153	1 MYMQN	myoglobin - dourou
35	690	86.9	153	1 MYGJ	myoglobin - common
36	690	86.9	153	1 MYBTF	myoglobin - Egypti
37	688	86.6	153	1 MYOT	myoglobin - Eurasi
38	688	86.6	153	1 MYLEM	myoglobin - weasel
39	687	86.5	153	1 MYOY	myoglobin - aardva
40	685	86.3	153	1 MYLR	myoglobin - slow l
41	684	86.1	153	1 MYZC	myoglobin - Califo
42	683	86.0	153	1 MYVTNG	myoglobin - northe
43	682	85.9	153	1 MYSLG	myoglobin - gray s
44	682	85.9	153	1 MYSLH	myoglobin [validat
45	679	85.5	153	1 MYGC	myoglobin - thick-

ALIGNMENTS

RESULT 1

MYWHP

myoglobin [validated] - sperm whale

C:Species: Physeter catodon (sperm whale)

C:Date: 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change 15-Sep-2000

C:Accession: A90591; A92844; A93150; A02506

R:Romero-Herrera, A.E.; Lehmann, H.

Biochim. Biophys. Acta 336, 318-323, 1974

A:Title: Residue 122 of sperm whale and horse myoglobin.

A:Reference number: A90591

A:Accession: A90591

A:Molecule type: protein

A:Residues: 1-153 <ROM>

A:Experimental source: skeletal muscle

R:Takano, T.

J. Mol. Biol. 110, 537-568, 1977

A:Title: Structure of myoglobin refined at 2.0 angstrom resolution. I. Crystallographic

A:Reference number: A92844; MUID:7714097; PMID:845959

A:Contents: X-ray crystallography of metmyoglobin, 2.0 angstroms

A:Accession: A92844

A:Molecule type: protein

A:Residues: 1-120,'A',122-153 <TAK>

R:Edmundson, A.B.

Nature 205, 883-887, 1965

A:Title: Amino-acid sequence of sperm whale myoglobin.

A:Reference number: A93150

A:Accession: A93150

A:Molecule type: protein

A:Residues: 1-121,'N',123-153 <EDM>

A:Experimental source: heart muscle

R:Watson, H.C.; Kendrew, J.C.

submitted to the Brookhaven Protein Data Bank, April 1973

A:Reference number: A50272; PDB:1MEN

A:Contents: annotation; X-ray crystallography, 2.0 angstroms, residues 1-153

R:Takano, T.

submitted to the Brookhaven Protein Data Bank, January 1988

A:Reference number: A50732; PDB:5MEN

A:Contents: annotation; X-ray crystallography, 2.0 angstroms, residues 1-153

R:Takano, T.

J. Mol. Biol. 110, 569-584, 1977

A:Title: Structure of myoglobin refined at 2.0 angstrom resolution. II. Structure of deo

A:Reference number: A92845; MUID:7714098; PMID:845960

A:Contents: annotation; X-ray crystallography of deoxymyoglobin, 2.0 angstroms

C:Function:

A:Description: binds molecular oxygen for intracellular storage and transport, primarily

C:Superfamily: globin; globin homology

C:Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier

F:2-147/Domain: globin homology <GLB>

F:64/Binding site: oxygen (His) (distal axial ligand) #status experimental

F:93/Binding site: heme iron (His) (proximal axial ligand) #status experimental

Query Match

100.0%; Score 794; DB 1; Length 153;

Best Local Similarity 100.0%; Pred. No. 8.4e-59;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASD 60
DB 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASD 60

QY 61 LKKGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120
DB 61 LKKGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120

QY 121 GDFGADAQAMNKALELFRKDIKAAKYKELGYQG 153
DB 121 GDFGADAQAMNKALELFRKDIKAAKYKELGYQG 153

RESULT 2

MYWHH

myoglobin - dwarf sperm whale
C:Species: Kogia simus (dwarf sperm whale)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 03-Mar-2000
C:Accession: A90407; A02506
R:Lehman, L.D.; Jones, B.N.; Lehman, L.D.; Gurd, F.R.N.
Biochemistry 16, 873-877, 1977
A:Title: The complete amino acid sequence of the major component myoglobin of dwarf sperm whale.
A:Reference number: A90407; MUID:77134684; PMID:843520
A:Accession: A90407
A:Molecule type: protein
A:Residues: 1-153 <DWU>
C:Superfamily: globin; globin homology
C:Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F:2-147/Domain: globin homology <GLB>
F:64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F:93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 97.4%; Score 773; DB 1; Length 153;
Best Local Similarity 96.7%; Pred. No. 4.6e-57;
Matches 148; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASD 60
DB 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASD 60

QY 61 LKKGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120
DB 61 LKKGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120

QY 121 GDFGADAQAMNKALELFRKDIKAAKYKELGYQG 153
DB 121 ADFGADAQAMNKALELFRKDIKAAKYKELGYQG 153

RESULT 3

MYWHC

myoglobin - California gray whale
C:Species: Eschrichtius robustus, Eschrichtius gibbosus (California gray whale)
C:Date: 01-Sep-1981 #sequence_revision 01-Sep-1981 #text_change 03-Mar-2000
C:Accession: A02499
R:Bogardt Jr., R.A.; Dwulet, F.E.; Jones, B.N.; Lehman, L.D.; Gurd, F.R.N.
Biochemistry 15, 2597-2602, 1976
A:Title: Complete primary structure of the major component myoglobin of California gray whale.
A:Reference number: A02499; MUID:76232191; PMID:938629
A:Accession: A02499
A:Molecule type: protein
A:Residues: 1-153 <BOG>
C:Comment: This myoglobin was isolated from skeletal muscle.
C:Superfamily: globin; globin homology
C:Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F:2-147/Domain: globin homology <GLB>
F:64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F:93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match

94.0%; Score 746; DB 1; Length 153;

Best Local Similarity 92.8%; Pred. No. 7.8e-55;
Matches 142; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASD 60
DB 1 VLSDAEWQLVLNIAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASD 60

QY 61 LKKGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120
DB 61 LKKGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120

QY 121 GDFGADAQAMNKALELFRKDIKAAKYKELGYQG 153
DB 121 GDFGADAQAMNKALELFRKDIKAAKYKELGYQG 153

RESULT 4

MYWHH

myoglobin - humpback whale
C:Species: Megaptera novaeangliae (humpback whale)
C:Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 03-Mar-2000
C:Accession: A90416; A91088; A02500
R:Lehman, L.D.; Dwulet, F.E.; Jones, B.N.; Bogardt Jr., R.A.; Krueckeberg, S.T.; Vissache Biochemistry 17, 3736-3739, 1978
A:Title: Complete amino acid sequence of the major component myoglobin from the humpback whale.
A:Reference number: A90416; MUID:79021601; PMID:698193
A:Accession: A90416
A:Molecule type: protein
A:Residues: 1-153 <LEH>
R:Edman, P.; Beggs, G.
Eur. J. Biochem. 1, 80-91, 1967
A:Title: A protein sequenator.
A:Reference number: A91088; MUID:68049168; PMID:6059350
A:Accession: A91088
A:Molecule type: protein
A:Residues: 1-60 <EDM>
A:Note: this was the first sequence determined using the modern sequenator
C:Comment: This myoglobin was isolated from skeletal muscle.
C:Superfamily: globin; globin homology
C:Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F:2-147/Domain: globin homology <GLB>
F:64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F:93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 93.2%; Score 740; DB 1; Length 153;
Best Local Similarity 92.2%; Pred. No. 2.5e-54;
Matches 141; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVLHWAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASD 60
DB 1 VLSDAEWQLVLNIAKVEADVAGHGQDILIRLFKSHPETLEKFDKFKHLKTEAEKASD 60

QY 61 LKKGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120
DB 61 LKKGVTVTALGAILKKKGHHEAELKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120

QY 121 GDFGADAQAMNKALELFRKDIKAAKYKELGYQG 153
DB 121 ADFGADAQAMNKALELFRKDIKAAKYKELGYQG 153

RESULT 5

MYDDAR

myoglobin - Amazon dolphin
C:Species: Inia geoffrensis (Amazon dolphin)
C:Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 03-Mar-2000
C:Accession: A02503
R:Dwulet, F.E.; Bogardt, R.A.; Jones, B.N.; Lehman, L.D.; Gurd, F.R.N.
Biochemistry 14, 5336-5343, 1975
A:Title: The complete amino acid sequence of the major component myoglobin of Amazon river dolphin.
A:Reference number: A02503; MUID:76062412; PMID:1191640
A:Accession: A02503
A:Molecule type: protein

A>Note: The saddleback dolphin is also called Black Sea dolphin
C>Date: 27-Nov-1985 #sequence revision 31-Mar-1991 #text_change 03-Mar-2000
C/Accession: A26230; A02497; A37509; A37510; A02494
R/Wang, C.C.; Avila, R.; Jones, B.N.; Gurd, F.R.N.
Biochemistry 15, 4978-4981, 1977

A>Title: Complete primary structure of the major component myoglobin of Pacific common dolphin
A/Reference number: A26230; MUID:178020893; PMID:911808
A/Accession: A26230
A/Molecule type: protein
A/Residues: 1-153 <WAN>
R/Kluh, I.: Bakardjieva, A.
FEBS Lett. 17, 31-34, 1971

A>Title: Primary structure of N-terminal part of molecule of dolphin myoglobin.
A/Reference number: A02497
A/Accession: A02497
A/Molecule type: protein
A/Residues: 1-20,'V',22-27,'I',29-31<KLU>
R/Bakardjieva, M.; Nedkov, P.; Bakardjieva, A.; Genov, N.
BIOCHIM Biophys. Acta 221, 136-139, 1970

A>Title: Difference in amino acid sequence between dolphin and sperm whale myoglobins.
A/Reference number: A37509; MUID:71014229; PMID:5473803
A/Accession: A37509
A/Molecule type: protein
A/Residues: 31'-65,'D','67'-121,'Q',123-153<KAR>
A>Note: The amidation states of residues 60, 85, and 132 were not determined and were as submitted to the Atlas, December 1977

R/Kluh, I.

A/Reference number: A37510
A/Contents: revision to residue 26
A/Accession: A37510
A/Molecule type: protein
A/Residues: 26<KL2>

C/Superfamily: globin; globin homology
C/Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F:2-147/Domain: globin homology <GLB>
F:64/Binding site: oxygen (His) [distal axial ligand] #status predicted
F:93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 92.3%; Score 733; DB 1; Length 153;
Best Local Similarity 90.8%; Pred.No. 9.3e-54;
Matches 138; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

Qy 2 LSEGEQLVLHWAKVEADVAGHGODILIRLFKSHPETLEKFDPKHLLTAEAMKASDDL 61
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 2 LSDGEGQLVLNVWGKVADIALGHGDVLIIRLFGHPETLEKFDPKHLLTADMKASDDL 61
 |||||
Qy 62 KHGGYTVLTALGAILLKKKGHEELPLAQASHATKKIKPIYLEFISEAIHVLHSRPG 121
 |||||
Db 62 KKHGNVTTLTGAILLKKGKHHDELPLAQASHATKKIKPIYLEFISEAIHVLHSRHPA 121
 |||||
Qy 122 DFGADAQQGMNKALELRFKDIAAKYKELGYYG 153
 :|||||
Db 122 EFGADAQQGMNKALELRFKDIAAKYKELGPHG 153
 ::|||:

RESULT 10
MYPE

myoglobin - harbor porpoise
C/Species: Phocoena phocaena (harbor porpoise)
C/Date: 27-Nov-1985 #sequence revision 27-Nov-1985 #text_change 03-Mar-2000
C/Accession: B92045; A90415; A02498
R/Bradshaw, R.A.; Gurd, F.R.N.
J. Biol. Chem. 244, 2167-2181, 1969

A>Title: Comparison of myoglobins from harbor seal, porpoise, and sperm whale.
A/Reference number: A92045; MUID:69177451; PMID:5782005
A/Accession: B92045
A/Molecule type: protein
A/Residues: 1-82,'EAN',86-153<BRA>
A>Note: This sequence has been revised in reference A90415
A>Note: This is the final paper in a series
R.Meuth, J.L.; Jones, B.N.; Garner, W.H.; Gurd, F.R.N.
Biochemistry 17, 3429-3431, 1978

Biochemistry 17, 3429-3431, 1978

A>Title: Complete amino acid sequence of the myoglobin from the Dall porpoise (Phocoenoides dallii)

A;Title: The complete amino acid sequence of the major component myoglobin from the Arch
A;Reference number: A02501; MUID:77112438; PMID:836810
A;Accession: A02501
A;Molecule type: protein
A;Residues: 1-153 <LEH>
C;Comment: This myoglobin was isolated from skeletal muscle.
C;Superfamily: globin; globin homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F;2-147/Domain: globin homology <GLB>
F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 92.1%; Score 731; DB 1; Length 153;
Best Local Similarity 90.8%; Pred. No. 1.4e-53;
Matches 139; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHVWAKVEADVAGHQDILIRLPKSHPTLEKFDKPKHLKTEAEKASEDL 60
Db 1 VLSDAEWHLVNIWAKVEADVAGHQDILIRLPKSHPTLEKFDKPKHLKTEAEKASEDL 60
QY 61 LKKGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
Db 61 LKKGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
QY 121 GDFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153
Db 121 ADFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153

RESULT 13
MYWHF
myoglobin - finback whale
C;Species: Balaeoptera physalus (finback whale, common roqual)
C;Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 03-Mar-2000
C;Accession: A02502
R;DiMarchi, R.D.; Wang, C.C.; Hemenway, J.B.; Gurd, F.R.N.
Biochemistry 17, 1968-1970, 1978
A;Title: Complete amino acid sequence of the major component myoglobin of finback whale
A;Reference number: A02502; MUID:78187212; PMID:656375
A;Accession: A02502
A;Molecule type: protein
A;Residues: 1-153 <DIM>
C;Comment: This myoglobin was isolated from skeletal muscle.
C;Superfamily: globin; globin homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F;2-147/Domain: globin homology <GLB>
F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 91.4%; Score 726; DB 1; Length 153;
Best Local Similarity 90.2%; Pred. No. 3.5e-53;
Matches 138; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHVWAKVEADVAGHQDILIRLPKSHPTLEKFDKPKHLKTEAEKASEDL 60
Db 1 VLSDAEWHLVNIWAKVEADVAGHQDILIRLPKSHPTLEKFDKPKHLKTEAEKASEDL 60
QY 61 LKKGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
Db 61 LKKGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
QY 121 GDFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153
Db 121 ADFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153

RESULT 14
MYWHU
myoglobin - Hubbs' whale
C;Species: Mesoplodon carlhubbsi (Hubbs' whale)
C;Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 03-Mar-2000
C;Accession: A02505
R;Dwulet, J.A.; Dwulet, F.E.; Gurd, F.R.N.

A;Title: The complete amino acid sequence of the major component myoglobin from the Arch
A;Reference number: A02501; MUID:77112438; PMID:836810
A;Accession: A02501
A;Molecule type: protein
A;Residues: 1-153 <LEH>
C;Comment: This myoglobin was isolated from skeletal muscle.
C;Superfamily: globin; globin homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F;2-147/Domain: globin homology <GLB>
F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 92.1%; Score 731; DB 1; Length 153;
Best Local Similarity 90.8%; Pred. No. 1.4e-53;
Matches 139; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHVWAKVEADVAGHQDILIRLPKSHPTLEKFDKPKHLKTEAEKASEDL 60
Db 1 VLSDAEWHLVNIWAKVEADVAGHQDILIRLPKSHPTLEKFDKPKHLKTEAEKASEDL 60
QY 61 LKKGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
Db 61 LKKGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
QY 121 GDFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153
Db 121 ADFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153

RESULT 13
MYWHF
myoglobin - finback whale
C;Species: Balaeoptera physalus (finback whale, common roqual)
C;Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 03-Mar-2000
C;Accession: A02502
R;DiMarchi, R.D.; Wang, C.C.; Hemenway, J.B.; Gurd, F.R.N.
Biochemistry 17, 1968-1970, 1978
A;Title: Complete amino acid sequence of the major component myoglobin of finback whale
A;Reference number: A02502; MUID:78187212; PMID:656375
A;Accession: A02502
A;Molecule type: protein
A;Residues: 1-153 <DIM>
C;Comment: This myoglobin was isolated from skeletal muscle.
C;Superfamily: globin; globin homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F;2-147/Domain: globin homology <GLB>
F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 91.4%; Score 726; DB 1; Length 153;
Best Local Similarity 90.2%; Pred. No. 3.5e-53;
Matches 138; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHVWAKVEADVAGHQDILIRLPKSHPTLEKFDKPKHLKTEAEKASEDL 60
Db 1 VLSDAEWHLVNIWAKVEADVAGHQDILIRLPKSHPTLEKFDKPKHLKTEAEKASEDL 60
QY 61 LKKGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
Db 61 LKKGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
QY 121 GDFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153
Db 121 ADFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153

RESULT 14
MYWHU
myoglobin - Hubbs' whale
C;Species: Mesoplodon carlhubbsi (Hubbs' whale)
C;Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 03-Mar-2000
C;Accession: A02505
R;Dwulet, J.A.; Dwulet, F.E.; Gurd, F.R.N.

Biochim. Biophys. Acta 624, 121-129, 1980
A;Title: Complete amino acid sequence of the major component myoglobin from Hubb's beake
A;Reference number: A02505; MUID:8100592; PMID:7407230
A;Accession: A02505
A;Molecule type: protein
A;Residues: 1-153 <DMU>
C;Comment: This myoglobin was isolated from skeletal muscle.
C;Superfamily: globin; globin homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F;2-147/Domain: globin homology <GLB>
F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 91.3%; Score 725; DB 1; Length 153;
Best Local Similarity 90.1%; Pred. No. 4.3e-53;
Matches 137; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHQDILIRLPKSHPTLEKFDKPKHLKTEAEKASEDL 61
Db 2 LSEGEWQLVHVWAKVEADVAGHQDILIRLPKSHPTLEKFDKPKHLKTEAEKASEDL 61
QY 62 KKHGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 121
Db 62 KKHGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 121
QY 122 DFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153
Db 122 DFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153

RESULT 15
MYWHZ
myoglobin - goose-beaked whale
C;Species: Ziphius cavirostris (goose-beaked whale)
C;Date: 01-Sep-1981 #sequence_revision 01-Sep-1981 #text_change 03-Mar-2000
C;Accession: A02504
R;Lehman, L.D.; Jones, B.N.; Dwulet, F.E.; Bogardt Jr., R.A.; Gurd, F.R.N.
Biochim. Biophys. Acta 625, 221-229, 1980
A;Title: Complete amino acid sequence of the major component myoglobin from the goose-be
A;Reference number: A02504; MUID:81063290; PMID:7437458
A;Accession: A02504
A;Molecule type: protein
A;Residues: 1-153 <LEH>
C;Comment: This myoglobin was isolated from skeletal muscle.
C;Superfamily: globin; globin homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; muscle; oxygen carrier
F;2-147/Domain: globin homology <GLB>
F;64/Binding site: oxygen (His) (distal axial ligand) #status predicted
F;93/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 90.9%; Score 722; DB 1; Length 153;
Best Local Similarity 90.1%; Pred. No. 7.6e-53;
Matches 137; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHQDILIRLPKSHPTLEKFDKPKHLKTEAEKASEDL 61
Db 2 LSEGEWQLVHVWAKVEADVAGHQDILIRLPKSHPTLEKFDKPKHLKTEAEKASEDL 61
QY 62 KKHGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 121
Db 62 KKHGHTVLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 121
QY 122 DFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153
Db 122 DFGADAQAAMNKALELFRKDIKAAKYKELGYQG 153

Search completed: August 10, 2004, 15:29:27
Job time : 7.66828 secs

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:19:00 ; Search time 4.81598 Seconds
(without alignments)
1654.229 Million cell updates/sec

Title: US-09-455-978b-76

Perfect score: 794

Sequence: 1 VLSGEWQLVHLVWAKVEAD.....ALELFRKDIAAKYKELGYQG 153

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	794	100.0	153	1	MYG_PHYCA	P02185 physeter ca
2	773	97.4	153	1	MYG_KOGSI	P02184 kogia simus
3	746	94.0	153	1	MYG_ESCGI	P02177 eschrichtiu
4	740	93.2	153	1	MYG_MEGNO	P02178 megaptera n
5	738	92.9	153	1	MYG_INIGE	P02181 inia geoffr
6	737	92.8	153	1	MYG_GLOME	P02174 globicephal
7	734	92.4	153	1	MYG_ORCOR	P02173 orcinus orc
8	733	92.3	153	1	MYG_TURTR	P02172 tursiops tr
9	731	92.1	153	1	MYG_BALAC	P02179 balaenopter
10	731	92.1	153	1	MYG_PROPH	P02176 phocoenoid
11	726	91.4	153	1	MYG_BALPH	P02180 balaenopter
12	725	91.3	153	1	MYG_MESCA	P02183 mesoplodon
13	722	90.9	153	1	MYG_CASFI	P14396 castor fibe
14	722	90.9	153	1	MYG_ZIPCA	P02182 ziphius cav
15	717	90.3	153	1	MYG_HORSE	P02188 equus cabal
16	707	89.0	153	1	MYG_PANTR	P02145 pan troglod
17	705	88.8	153	1	MYG_PERPO	P02166 perodicticu
18	703	88.5	153	1	MYG_RABIT	P02170 oryctolagus
19	701	88.3	153	1	MYG_GORBE	P02147 gorilla gor
20	701	88.3	153	1	MYG_ONDZI	P32428 ondatra zib
21	700	88.2	153	1	MYG_PIG	P02189 sus scrofa
22	699	88.0	153	1	MYG_HUMAN	P02144 homo sapien
23	695	87.5	153	1	MYG_MACFA	P02150 macaca fasc
24	694	87.4	153	1	MYG_PONPY	P02148 pongo pygma
25	693	87.3	153	1	MYG_HYLAG	P02146 hylobates a
26	693	87.3	153	1	MYG_TUPLG	P02165 tupiaia glis
27	691	87.0	153	1	MYG_AOTTR	P02151 aotus trivi
28	691	87.0	153	1	MYG_PAPAN	P02149 papio anubi
29	690	86.9	153	1	MYG_CALJA	P02152 callithrix
30	690	86.9	153	1	MYG_ROUAE	P02163 rousettus a
31	688	86.6	153	1	MYG_LEPMU	P02169 lepitemur m
32	688	86.6	153	1	MYG_LUTLU	P11343 lutra lutra
33	687	86.5	153	1	MYG_ORYAF	P02164 orycteropus

34	685	86.3	153	1	MYG_NYCCO	P02167 nycticebus
35	684	86.1	153	1	MYG_ZALCA	P02161 zalophus ca
36	683	86.0	153	1	MYG_CTEGU	P20856 ctenodactyl
37	682	85.9	153	1	MYG_HALGR	P02162 halichoerus
38	680	85.6	153	1	MYG_DELLE	P83682 delphinapte
39	679	85.5	153	1	MYG_GALCR	P02168 galago cras
40	678	85.4	153	1	MYG_CEBAP	P02153 cebus apell
41	678	85.4	153	1	MYG_LAGLA	P02154 lagothrix l
42	678	85.4	153	1	MYG_OCHPR	P02171 ochotona pr
43	677	85.3	153	1	MYG_SPAEH	P04248 spalax leuc
44	674	84.9	153	1	MYG_CEREL	P02191 cervus elap
45	674	84.9	153	1	MYG_PROGU	P04249 proechimys

ALIGNMENTS

RESULT 1
MYG_PHYCA
AC P02185; STANDARD; PRT; 153 AA.
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myoglobin.
GN MB.
OS Physeter catodon (Sperm whale) (Physeter macrocephalus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti;
OC Physetridae; Physeter.
OX NCBI_TaxID=9755;
RN [1]
RP SEQUENCE.
RC TISSUE=Heart muscle;
RA Edmundson A.B.;
RT "Amino-acid sequence of sperm whale myoglobin.";
RL Nature 205:883-887(1965).
RN [2]
RP REVISION TO 122.
RC TISSUE=Skeletal muscle;
RA Romero-Herrera A.E., Lehmann H.;
RT "Residue 122 of sperm whale and horse myoglobin.";
RL Biochim. Biophys. Acta 336:318-323(1974).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF METMYOGLOBIN.
RX MEDLINE=77144097; PubMed=845959;
RA Takano T.;
RT "Structure of myoglobin refined at 2.0-A resolution. I.
J. Mol. Biol. 110:537-568(1977).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF DEOXYMYOGLOBIN.
RX MEDLINE=77144098; PubMed=845960;
RA Takano T.;
RT "Structure of myoglobin refined at 2.0-A resolution. II. Structure of
deoxymyoglobin from sperm whale.";
J. Mol. Biol. 110:569-584(1977).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF OXYMYOGLOBIN.
RX MEDLINE=81119812; PubMed=7463482;
RA Phillips S.E.V.;
RT "Structure and refinement of oxymyoglobin at 1.6-A resolution.";
J. Mol. Biol. 142:531-554(1980).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=91132649; PubMed=1994031;
RA Lionetti C., Guanziroli M.G., Frigerio F., Ascenzi P., Bolognesi M.;
RT "X-ray crystal structure of the ferric sperm whale myoglobin:
imidazole complex at 2.0-A resolution.";
J. Mol. Biol. 217:409-412(1991).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX MEDLINE=99158603; PubMed=10049310;

RA Brunori M., Cutruzzola F., Savino C., Travaglini-Allocatelli C.,
RA Vallone B., Gibson Q.H.;
RT "Structural dynamics of ligand diffusion in the protein matrix: a
RT study on a new myoglobin mutant Y(B10) Q(E7) R(B10).";
RL Biophys. J. 76:1259-1269(1999).
RN [8].
RP X-RAY CRYSTALLOGRAPHY (1.4 ANGSTROMS).
RX MEDLINE=20160904; PubMed=10681426;
RA Brunori M., Vallone B., Cutruzzola F., Travaglini-Allocatelli C.,
RA Berendzen J., Chu K., Sweet R.M., Schlichting I.;
RT "The role of cavities in protein dynamics: crystal structure of a
RT photocyclic intermediate of a mutant myoglobin";
RL Proc. Natl. Acad. Sci. U.S.A. 97:2058-2063(2000).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates
CC the movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
DR PIR; A90591; MYWP.
DR PDB; 101M; 08-APR-98.
DR PDB; 102M; 08-APR-98.
DR PDB; 103M; 08-APR-98.
DR PDB; 104M; 08-APR-98.
DR PDB; 105M; 08-APR-98.
DR PDB; 106M; 08-APR-98.
DR PDB; 107M; 08-APR-98.
DR PDB; 108M; 20-MAY-98.
DR PDB; 109M; 08-APR-98.
DR PDB; 110M; 08-APR-98.
DR PDB; 111M; 08-APR-98.
DR PDB; 112M; 08-APR-98.
DR PDB; 1A6G; 21-OCT-98.
DR PDB; 1A6K; 06-APR-99.
DR PDB; 1A6M; 06-APR-99.
DR PDB; 1A6N; 06-APR-99.
DR PDB; 1A6S; 01-APR-97.
DR PDB; 1AJG; 12-NOV-97.
DR PDB; 1AJH; 12-NOV-97.
DR PDB; 1BVC; 31-JUL-95.
DR PDB; 1BVD; 31-JUL-95.
DR PDB; 1BZ6; 11-NOV-98.
DR PDB; 1BZP; 10-MAY-99.
DR PDB; 1BZR; 10-MAY-99.
DR PDB; 1CH1; 09-APR-99.
DR PDB; 1CH2; 09-APR-99.
DR PDB; 1CH3; 09-APR-99.
DR PDB; 1CH5; 09-APR-99.
DR PDB; 1CH7; 09-APR-99.
DR PDB; 1CH9; 09-APR-99.
DR PDB; 1CIK; 09-APR-99.
DR PDB; 1CIO; 09-APR-99.
DR PDB; 1CO8; 14-JUN-99.
DR PDB; 1CO9; 14-JUN-99.
DR PDB; 1CF0; 14-JUN-99.
DR PDB; 1CF5; 14-JUN-99.
DR PDB; 1CFW; 14-JUN-99.
DR PDB; 1CO2; 16-AUG-99.
DR PDB; 1DO1; 02-APR-00.
DR PDB; 1DO3; 02-APR-00.
DR PDB; 1DO4; 02-APR-00.
DR PDB; 1DO7; 02-APR-00.
DR PDB; 1DTI; 30-JAN-00.
DR PDB; 1DTM; 14-JUN-00.
DR PDB; 1DUK; 08-APR-03.
DR PDB; 1DOO; 27-DEC-00.
DR PDB; 1DXC; 02-APR-00.
DR PDB; 1DXD; 20-JUL-00.
DR PDB; 1EBC; 13-AUG-99.
DR PDB; 1F63; 19-JUL-00.
DR PDB; 1F65; 19-JUL-00.
DR PDB; 1F6H; 21-JUL-00.
DR PDB; 1FCS; 31-OCT-93.
DR PDB; 1HJT; 12-NOV-97.
DR PDB; 1IOP; 08-APR-98.
DR PDB; 1IRC; 11-JUL-96.
DR PDB; 1JDO; 27-MAY-98.
DR PDB; 1JF6; 16-JAN-02.
DR PDB; 1JP8; 16-JAN-02.
DR PDB; 1JP9; 16-JAN-02.
DR PDB; 1JPB; 16-JAN-02.
DR PDB; 1JW8; 10-OCT-01.
DR PDB; 1L2K; 21-AUG-02.
DR PDB; 1LTW; 23-DEC-96.
DR PDB; 1MBC; 09-JAN-89.
DR PDB; 1MBD; 27-OCT-83.
DR PDB; 1MBI; 15-OCT-91.
DR PDB; 1MBN; 27-OCT-83.
DR PDB; 1MBO; 30-SEP-83.
DR PDB; 1MCI; 07-DEC-95.
DR PDB; 1MGN; 31-JAN-94.
DR PDB; 1MLF; 31-AUG-94.
DR PDB; 1MLG; 31-AUG-94.
DR PDB; 1MLH; 31-AUG-94.
DR PDB; 1MLJ; 31-AUG-94.
DR PDB; 1MLK; 31-AUG-94.
DR PDB; 1MLL; 31-AUG-94.
DR PDB; 1MLM; 31-AUG-94.
DR PDB; 1MLN; 31-AUG-94.
DR PDB; 1MLO; 31-AUG-94.
DR PDB; 1MLQ; 07-FEB-95.
DR PDB; 1MLR; 31-AUG-94.
DR PDB; 1MLS; 31-AUG-94.
DR PDB; 1MLU; 31-AUG-94.
DR PDB; 1MOA; 07-FEB-95.
DR PDB; 1MOB; 07-FEB-95.
DR PDB; 1MOC; 07-FEB-95.
DR PDB; 1MOD; 07-FEB-95.
DR PDB; 1MTI; 15-SEP-95.
DR PDB; 1MTJ; 15-SEP-95.
DR PDB; 1MTK; 15-SEP-95.
DR PDB; 1MYF; 27-FEB-95.
DR PDB; 1MYM; 31-JAN-94.
DR PDB; 1OBM; 08-APR-98.
DR PDB; 1OFJ; 27-MAY-98.
DR PDB; 1OFK; 11-NOV-98.
DR PDB; 1SPE; 08-MAR-96.
DR PDB; 1SWM; 31-JAN-94.
DR PDB; 1TES; 08-NOV-96.
DR PDB; 1VXA; 01-AUG-96.
DR PDB; 1VXB; 01-AUG-96.
DR PDB; 1VXC; 01-AUG-96.
DR PDB; 1VXE; 01-AUG-96.
DR PDB; 1VXF; 01-AUG-96.
DR PDB; 1VXG; 01-AUG-96.
DR PDB; 1VXH; 01-AUG-96.
DR PDB; 1YOG; 07-DEC-96.
DR PDB; 1YOH; 07-DEC-96.
DR PDB; 1YOI; 07-DEC-96.
DR PDB; 2CMM; 31-JAN-94.
DR PDB; 2MB5; 15-APR-91.
DR PDB; 2MBW; 23-DEC-96.
DR PDB; 2MGA; 31-JAN-94.
DR PDB; 2MGB; 31-JAN-94.
DR PDB; 2MGC; 31-JUL-94.
DR PDB; 2MGD; 31-JUL-94.
DR PDB; 2MGE; 31-JUL-94.
DR PDB; 2MGF; 31-JUL-94.
DR PDB; 2MGG; 31-JUL-94.
DR PDB; 2MGH; 31-JUL-94.
DR PDB; 2MGI; 31-JUL-94.
DR PDB; 2MGJ; 31-JUL-94.
DR PDB; 2MGK; 31-JUL-94.
DR PDB; 2MGL; 31-JUL-94.
DR PDB; 2MGW; 31-JUL-94.
DR PDB; 2MYA; 31-JAN-94.
DR PDB; 2MYB; 31-JAN-94.
DR PDB; 2MYC; 31-JAN-94.

```

Query Match          100.0%; Score 794; DB 1; Length 153;
Best Local Similarity 100.0%; Pred. No. 1.2e-61;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLSGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEWKASED 60
DB 1 VLSGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEWKASED 60

QY 61 LKKGVTVTALGAILKKKGHEAEKPLAQSHATKHKIPKYLEFISEAIIHVLSRHP 120
DB 61 LKKGVTVTALGAILKKKGHEAEKPLAQSHATKHKIPKYLEFISEAIIHVLSRHP 120

QY 121 GDFGADAQAMNKALELFRKDIAAKYKELGYQG 153
DB 121 GDFGADAQAMNKALELFRKDIAAKYKELGYQG 153

RESULT 2
MYG_KOGSI
ID MYG_KOGSI STANDARD; PRT; 153 AA.
AC P02184;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myoglobin.
GN MB.
OS Kogia simus (Dwarf sperm whale).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti;
OC Physeteridae; Kogia.
OX NCBI_TaxID=9752;
RN [1]
RP SEQUENCE.
RX MEDLINE=77134684; PubMed=843520;
RA Dwulet F.E., Jones B.N., Lehman L.D., Gurd F.R.N.;
RT "The complete amino acid sequence of the major component myoglobin of
  dwarf sperm whale (Kogia simus).";
RL Biochemistry 16:873-877(1977).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
  movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
DR PIR; A90407; MYWHW.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND).
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
FT METAL 94 94 IRON (HEME DISTAL LIGAND).
SQ SEQUENCE 153 AA; 17237 MW; 5771A432C7B32614 CRC64;

Query Match          97.4%; Score 773; DB 1; Length 153;
Best Local Similarity 96.7%; Pred. No. 7.5e-60;
Matches 148; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VLSGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEWKASED 60
DB 1 VLSGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEWKASED 60

QY 61 LKKGVTVTALGAILKKKGHEAEKPLAQSHATKHKIPKYLEFISEAIIHVLSRHP 120
DB 61 LKKGVTVTALGAILKKKGHEAEKPLAQSHATKHKIPKYLEFISEAIIHVLSRHP 120

QY 121 GDFGADAQAMNKALELFRKDIAAKYKELGYQG 153
DB 121 ADFGADAQAMNKALELFRKDIAAKYKELGYQG 153

RESULT 3
MYG_ESCGI
ID MYG_ESCGI STANDARD; PRT; 153 AA.

```

```

AC P02177;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myoglobin.
GN MB.
OS Eschrichtius gibbosus (California gray whale) (Eschrichtius robustus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
OC Eschrichtiidae; Eschrichtius.
OX NCBI_TaxID=9764;
RN [1]
RP SEQUENCE.
RX TISSUE=Skeletal muscle;
RX MEDLINE=76232191; PubMed=938629;
RA Bogardt R.A. Jr., Dwulet F.E., Lehman L.D., Jones B.N., Gurd F.R.N.;
RT "Complete primary structure of the major component myoglobin of
  California gray whale (Eschrichtius gibbosus).";
RL Biochemistry 15:2597-2602(1976).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
  movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
DR PIR; A02499; MYWHC.
DR HSSP; P02185; 1B26.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND).
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
FT METAL 94 94 IRON (HEME DISTAL LIGAND).
SQ SEQUENCE 153 AA; 17118 MW; 4A5587BA0E805C7D CRC64;

Query Match          94.0%; Score 746; DB 1; Length 153;
Best Local Similarity 92.8%; Pred. No. 1.6e-57;
Matches 142; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 VLSGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEWKASED 60
DB 1 VLSDAEWQLVNIWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEWKASED 60

QY 61 LKKGVTVTALGAILKKKGHEAEKPLAQSHATKHKIPKYLEFISEAIIHVLSRHP 120
DB 61 LKKGVTVTALGAILKKKGHEAEKPLAQSHATKHKIPKYLEFISEAIIHVLSRHP 120

QY 121 GDFGADAQAMNKALELFRKDIAAKYKELGYQG 153
DB 121 GDFGADAQAMNKALELFRKDIAAKYKELGYQG 153

RESULT 4
MYG_MEGNO
ID MYG_MEGNO STANDARD; PRT; 153 AA.
AC P02178;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myoglobin.
GN MB.
OS Megaptera novaeangliae (Humpback whale).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
OC Balaenopteridae; Megaptera.
OX NCBI_TaxID=9773;
RN [1]
RP SEQUENCE.
RX TISSUE=Skeletal muscle;
RX MEDLINE=79021601; PubMed=698193;
RA Lehman L.D., Dwulet F.E., Jones B.N., Bogardt R.A. Jr.,
  Krueckeberg S.T., Visscher R.B., Gurd F.R.N.;
RT "Complete amino acid sequence of the major component myoglobin from
  the humpback whale, Megaptera novaeangliae.";

```

```

RL Biochemistry 17:3736-3739 (1978).
RN [2]
RP SEQUENCE OF 1-60.
RX MEDLINE=68049168; PubMed=6059350;
RA Edman P., Begg G.;
RT "A protein sequenator.";
RL Eur. J. Biochem. 1:80-91(1967).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
CC movement of oxygen within muscles.
CC -!- MISCELLANEOUS: This sequence was the first determined using the
CC Edman sequencing system.
CC -!- SIMILARITY: Belongs to the globin family.
DR PIR; A90416; MYWHH.
DR HSSP; P02185; 1B26.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; Globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND)..
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 153 AA; 17132 MW; 4FAFD81A0B805C7D CRC64;

Query Match 93.2%; Score 740; DB 1; Length 153;
Best Local Similarity 92.2%; Pred. No. 5.2e-57;
Matches 141; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 VLSEGEWQLVHVWAKVADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKASD 60
DB 1 VLSDAEWQLVNIWAKVADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKASD 60

QY 61 LKKGVTVTALGAILKKKGHEAEKPLAQSHATKHKPIKYLEFISEAIHVLHSRHP 120
DB 61 LKKGNTVTALGGILKKKGHEAEKPLAQSHATKHKPIKYLEFISEAIHVLHSRHP 120

QY 121 GDFGADAQGMNKALELFRKIDIAAKYKELGYQG 153
DB 121 ADFGADAQGMNKALELFRKIDIAAKYKELGFGQ 153

RESULT 5
ID MYG INICE STANDARD; PRT; 153 AA.
AC P02181;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myoglobin.
GN MB.
OS Inia geoffrensis (Amazon dolphin).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Iniidae;
OC Inia.
OX NCBI_TaxID=9725;
RN [1]_
RP SEQUENCE.
RC TISSUE=Skeletal muscle;
RX MEDLINE=76062412; PubMed=1191640;
RA Dwulet F.E., Bogardt R.A., Jones B.N., Lehman L.D., Gurd F.R.N.;
RT "The complete amino acid sequence of the major component myoglobin of
RT Amazon river dolphin (Inia geoffrensis).";
RL Biochemistry 14:5336-5343(1975).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
CC movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
DR PIR; A02503; MYDDAR.
DR HSSP; P02188; 1WLA.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; Globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.

```

```

KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND)..
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 153 AA; 17071 MW; 72222EDD4EBCE5D CRC64;

Query Match 92.9%; Score 738; DB 1; Length 153;
Best Local Similarity 91.4%; Pred. No. 7.7e-57;
Matches 139; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKASD 61
DB 2 LSDGEWQLVNIWKGVEADLAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKASD 61

QY 62 KKHGVTVTALGAILKKKGHEAEKPLAQSHATKHKPIKYLEFISEAIHVLHSRHP 121
DB 62 KKHGNTVTALGGILKKKGHEAEKPLAQSHATKHKPIKYLEFISEAIHVLHSRHP 121

QY 122 DFGADAQGMNKALELFRKIDIAAKYKELGYQG 153
DB 122 DFGADAQGMNKALELFRKIDIAAKYKELGFGH 153

RESULT 6
ID MYG GLOME STANDARD; PRT; 153 AA.
AC P02174;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myoglobin.
GN MB.
OS Globicephala melaleuca (Long-finned pilot whale) (Globicephala melaleuca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;
OC Globicephala.
OX NCBI_TaxID=9731;
RN [1]_
RP SEQUENCE.
RC TISSUE=Skeletal muscle;
RX MEDLINE=78187213; PubMed=656376;
RA Jones B.N., Dwulet F.E., Lehman L.D., Garner M.H., Bogardt R.A. Jr.,
RA Garner W.H., Gurd F.R.N.;
RT "Complete amino acid sequence of myoglobin from the pilot whale,
RT Globicephala melaleuca.";
RL Biochemistry 17:1971-1974(1978).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
CC movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
DR PIR; A02496; MYWHT.
DR HSSP; P02185; 1A6W.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; Globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND)..
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 153 AA; 17085 MW; 510980DECB83DFD CRC64;

Query Match 92.8%; Score 737; DB 1; Length 153;
Best Local Similarity 92.1%; Pred. No. 9.4e-57;
Matches 140; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVADVAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKASD 61
DB 2 LSDGEWQLVNIWKGVEADLAGHGQDILIRLFKSHPETLEKFDPRFKHLKTEAEKASD 61

QY 62 KKHGVTVTALGAILKKKGHEAEKPLAQSHATKHKPIKYLEFISEAIHVLHSRHP 121
DB 62 KKHGNTVTALGGILKKKGHEAEKPLAQSHATKHKPIKYLEFISEAIHVLHSRHP 121

QY 122 DFGADAQGMNKALELFRKIDIAAKYKELGYQG 153

```



```

Db      122 EFGADAQAGMNAKLELFRKDIAAKYKELGFGH 153
          :|||||:|||||:|||||:|||||:|||||:|

RESULT 7
MYG_ORCOR MYG_ORCOR STANDARD; PRT; 153 AA.
ID_P02I73;
AC 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myoglobin..
GN MB..
OS Orcinus orca (Killer whale).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;
OC Orcinus.
OX NCBI_TaxID=9733;
RN [1]
RP SEQUENCE.
RC TISSUE=Skeletal muscle;
RX MEDLINE=81267443; PubMed=6115067;
RA Meuth J.L., Jones B.N., Gard F.R.N.;
RT Reassignment of residue 122 in the myoglobin from the killer whale,
RL "Orcinus orca.";
RN J. Mol. Evol. 17:163-166(1981).
[2]
RP SEQUENCE.
RC TISSUE=Skeletal muscle;
RX MEDLINE=77134898; PubMed=849459;
RA Castillo O., Lehmann H., Jones L.T.;
RT "The myoglobin of the killer whale (Orcinus orca).";
RL Biochim. Biophys. Acta 491:23-28(1977).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
CC movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
PIR; A92956; MYWHL.
DR HSPP; P02185; 1A6M.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64
FT FTAL 93 93
FT CONFLICT 122 122 E -> Q (IN REF. 2).
FT SEQUENCE 153 AA; 17071 MW; A408C49BCBB93C4C CRC64;

Query Match 92.4%; Score 734; DB 1; Length 153;
Best Local Similarity 91.4%; Pred. No. 1.7e-56;
Matches 139; Conservative 8; Mismatches 5; Indels 0; Gaps 0

Qy 2 LSEGEWQLVLHWVKAEDVAGHGDDIIIRLPKSHPETLEKPDPRFKHLKTEAEWKASEDL 61
Dd 2 LSDGEWQLVLNWGVKEADLAGHGDDIIIRLFKGHPETLEKFPDKPKHLKTEADMKASEDL 61
Qy 62 KKHGVTVLTALGAILKKKGHHAEALKPLAQSHATKHKIPIKYLEFISEAIHVLSHRHPG 121
Dd 62 KKHGVTVLTALGAILKKKGHHDAEALKPLAQSHATKHKIPIKYLEFISEAIHVLSHRHPA 121
Qy 122 EFGADAQAGMNAKLELFRKDIAAKYKELGYQG 153
Dd 122 EFGADAQAGMNAKLELFRKDIAAKYKELGFGH 153

RESULT 8
MYG_TURTR MYG_TURTR STANDARD; PRT; 153 AA.
ID_P02I72;
AC 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)

```

QY 62 KKHGVTTLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPG 121
 Db 62 KKHGNTVLTALGAILKKKGHDAELKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPA 121
 QY 122 DFGADAQAMNKALELFRKDIKAAKYKELGYOG 153
 Db 122 BFGADAQAMNKALELFRKDIKAAKYKELGFGH 153

RESULT 9

MYG_BALAC
 ID MYG_BALAC STANDARD; PRT; 153 AA.
 AC P02179;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myoglobin.
 GN MB.
 OS Balaenoptera acutorostrata (Minke whale) (Lesser rorqual).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
 OC Balaenopteridae; Balaenoptera.
 OX NCBI_TaxID=9767;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=77112438; PubMed=836810;
 RA Lehman L.D., Dwiulet F.E., Bogardt R.A. Jr., Jones B.N., Gurd F.R.N.;
 RT "The complete amino acid sequence of the major component myoglobin
 from the arctic minke whale, Balaenoptera acutorostrata.";
 RL Biochemistry 16:706-709(1977).
 CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
 movement of oxygen within muscles.
 CC -!- SIMILARITY: Belongs to the globin family.
 DR PIR; A02501; MYWK.
 DR HSSP; P02185; 1B26.
 DR InterPro; IPR000971; Globin.
 DR InterPro; IPR002335; Myoglobin.
 DR Pfam; PF00042; Globin; 1.
 DR PRINTS; PR00613; MYOGLOBIN.
 DR PROSITE; PS01033; GLOBIN; 1.
 KW Heme; Oxygen transport; Transport; Muscle.
 FT METAL 64 64 IRON (HEME DISTAL LIGAND).
 FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
 SQ SEQUENCE 153 AA; 17155 MW; A5364E71B9705C6E CRC64;

Query Match 92.1%; Score 731; DB 1; Length 153;
 Best Local Similarity 90.8%; Pred. No. 3.1e-56;
 Matches 139; Conservative 7; Mismatches 7; Indels 0; Gaps 0;
 QY 1 VLSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKHKTAEAMKASED 60
 Db 1 VLSDAEHLVHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKHKTAEAMKASED 60
 QY 61 LKHGVTTLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHP 120
 Db 61 LKHGNTVLTALGGILKKKGHHEAEKPLAQSHATKHKIPKYLEFISDAIHVLHSRHP 120
 QY 121 DFGADAQAMNKALELFRKDIKAAKYKELGYOG 153
 Db 121 AEFGADAQAMNKALELFRKDIKAAKYKELGFGH 153

RESULT 10

MYG_PHOPH
 ID MYG_PHOPH STANDARD; PRT; 153 AA.
 AC P02176;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myoglobin.
 GN MB.
 OS Phocoenoides phocoena (Harbor porpoise), and

OS Phocoenoides dalli dalli (Dall's porpoise).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Phocoenidae;
 OC Phocoena.
 OX NCBI_TaxID=9742, 9745;
 RN [1]
 RP SEQUENCE.
 RC SPECIES=P. phocoena;
 RX MEDLINE=69177451; PubMed=5782005;
 RA Bradshaw R.A., Gurd F.R.N.;
 RT "Comparison of myoglobins from harbor seal, porpoise, and sperm
 whale. V. The complete amino acid sequences of harbor seal and
 porpoise myoglobins.";
 RL J. Biol. Chem. 244:2167-2181(1969).
 RN [2]
 RP SEQUENCE, AND REVISIONS TO 83 AND 85.
 RC SPECIES=P. phocoena, and P.D.dalli;
 RX MEDLINE=79000346; PubMed=687594;
 RA Meuth J.L., Jones B.N., Garner W.H., Gurd F.R.N.;
 RT "Complete amino acid sequence of the myoglobin from the Dall porpoise
 (Phocoenoides dalli dalli) and reinvestigation of the primary
 structure of the myoglobin from common porpoise (Phocoena
 phocoena).";
 RL Biochemistry 17:3429-3431(1978).
 CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
 movement of oxygen within muscles.
 CC -!- SIMILARITY: Belongs to the globin family.
 DR PIR; B90415; MYPD.
 DR PIR; B92045; MYPE.
 DR HSSP; P02185; 1B26.
 DR InterPro; IPR000971; Globin.
 DR InterPro; IPR002335; Myoglobin.
 DR Pfam; PF00042; Globin; 1.
 DR PRINTS; PR00613; MYOGLOBIN.
 DR PROSITE; PS01033; GLOBIN; 1.
 KW Heme; Oxygen transport; Transport; Muscle.
 FT METAL 64 64 IRON (HEME DISTAL LIGAND).
 FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
 SQ SEQUENCE 153 AA; 17101 MW; 3FA2F4561A35CC2E CRC64;
 Query Match 92.1%; Score 731; DB 1; Length 153;
 Best Local Similarity 90.8%; Pred. No. 3.1e-56;
 Matches 138; Conservative 7; Mismatches 7; Indels 0; Gaps 0;
 QY 2 LSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKHKTAEAMKASEDL 61
 Db 2 LSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRKHKTAEAMKASEDL 61
 QY 62 KKHGVTTLTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPG 121
 Db 62 KKHGNTVLTALGGILKKKGHDAELKPLAQSHATKHKIPKYLEFISEAIHVLHSRHPA 121
 QY 122 DFGADAQAMNKALELFRKDIKAAKYKELGYOG 153
 Db 122 BFGADAQAMNKALELFRKDIKATKYKELGFGH 153
 RESULT 11
 MYG_BALPH
 ID MYG_BALPH STANDARD; PRT; 153 AA.
 AC P02180;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myoglobin.
 GN MB.
 OS Balaenoptera physalus (Finback whale) (Common rorqual).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
 OC Balaenopteridae; Balaenoptera.
 OX NCBI_TaxID=9770;
 RN [1]
 RP SEQUENCE.

```

RC TISSUE=Skeletal muscle;
RX MEDLINE=78187212; PubMed=656375;
RA Dimarchi R.D., Wang C.-C., Hemenway J.B., Gurd F.R.N.;
RT "Complete amino acid sequence of the major component myoglobin of
RL finback whale (Balaenoptera physalus).";
RL Biochemistry 17:1968-1970(1978).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
CC movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
CC PIR; A02502; MYWHF.
DR HSSP; P02185; 1B26.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND).
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 153 AA; 17086 MW; 8AFC89C1BF00291F CRC64;

Query Match 91.4%; Score 726; DB 1; Length 153;
Best Local Similarity 90.2%; Pred. No. 8.4e-56;
Matches 138; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 VLSGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFRFKHLKTEAEWKASED 60
DB 1 VLTDAEHLVLINWAKVEADVAGHGQDILISLFKGPETLEKFRFKHLKTEAEWKASED 60
QY 61 LKHGVTVTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120
DB 61 LKHGNTVLTALGGILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 120
QY 121 GDFGADAQAMNKALELFRKDIKAAKYKELGYQG 153
DB 121 ADFGADAQAMNKALELFRKDIKAAKYKELGFQG 153

RESULT 12
MYG_MESCA
ID MYG_MESCA STANDARD; PRT; 153 AA.
AC P02183;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myoglobin.
GN MB.
OS Mesopodion carlhubbsi (Hubb's whale).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Ziphiidae;
OC Mesopodion.
OX NCBI_taxID=9758;
RN [1]
RP SEQUENCE.
RC TISSUE=Skeletal muscle;
RX MEDLINE=81000592; PubMed=7407230;
RA Dwalet J.A., Dwalet F.E., Gurd F.R.N.;
RT "Complete amino acid sequence of the major component myoglobin from
RT Hubb's beaked whale, Mesopodion carlhubbsi.";
RL Biochim. Biophys. Acta 624:121-129(1980).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
CC movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
CC PIR; A02505; MYWHU.
DR HSSP; P02185; 1A6M.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND).
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).

```

```

SQ SEQUENCE 153 AA; 17136 MW; 8BA3DB535CBE91B9 CRC64;

Query Match 91.3%; Score 725; DB 1; Length 153;
Best Local Similarity 90.1%; Pred. No. 1e-55;
Matches 137; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFRFKHLKTEAEWKASEDL 61
DB 2 LSEABWQLVHVWAKVEADLSGHQGEILIRLFKGPETLEKFRFKHLKTEAEWKASEDL 61
QY 62 KKHGVTVTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 121
DB 62 KKHGHTVLTALGGILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIIHVLHSKHP 121
QY 122 DFGADAQAMNKALELFRKDIKAAKYKELGYQG 153
DB 122 DFGADAQAMTKALELFRKDIKAAKYKELGFPG 153

RESULT 13
MYG_CASFI
ID MYG_CASFI STANDARD; PRT; 153 AA.
AC P14396;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myoglobin.
GN MB.
OS Castor fiber (Eurasian beaver).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Castoridae; Castor.
OX NCBI_taxID=10185;
RN [1]
RP SEQUENCE.
RA Sukhomlinov B.F., Drobot L.B.;
RT "The primary structure of myoglobin from beaver (Castor fiber); II
RT peptic peptides of the tryptic insoluble core. Reconstruction of
RT beaver myoglobin polypeptide chain.";
RL Bioorg. Khim. 6:542-546(1980).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
CC movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
DR PIR; JN0410; JN0410.
DR HSSP; P02189; 1MWC.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND).
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 153 AA; 17020 MW; 4FD93C4E116B6D4D CRC64;

Query Match 90.9%; Score 722; DB 1; Length 153;
Best Local Similarity 87.5%; Pred. No. 1.8e-55;
Matches 133; Conservative 13; Mismatches 6; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKFRFKHLKTEAEWKASEDL 61
DB 2 LSDGEWQLVHVWGKVEADLAGHQEVILIRLFKGPETLEKFNKFKHKSDEMKASEDL 61
QY 62 KKHGVTVTALGAILKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIIHVLHSRHP 121
DB 62 KKHGVTVTALGGVLKKKGHHEAEKPLAQSHATKHKIPKYLEFISEAIIHVLQSKHP 121
QY 122 DFGADAQAMNKALELFRKDIKAAKYKELGYQG 153
DB 122 BFGADABGAMNKALELFRKDIKAAKYKELGFQG 153

RESULT 14
MYG_ZIPCA

```

```

ID MYG_ZIPCA STANDARD; PRT; 153 AA.
AC P02182;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myoglobin.
GN MB.
OS Ziphium cavirostris (Goose-beaked whale).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Ziphiidae;
OC Ziphium.
OX NCBI_TaxID=9760;
RN [1]_
RP SEQUENCE.
RX MEDLINE=81063290; PubMed=7437458;
RA Lehman L.D., Jones B.N., Dwulet F.E., Bogardt R.A. Jr., Gurd F.R.N.;
RT "Complete amino acid sequence of the major component myoglobin from
the goose-beaked whale, Ziphium cavirostris.";
RL Biochim. Biophys. Acta 625:221-229(1980).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
DR PIR; A02504; MYWZ.
DR HSP; P02185; IAGM.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; Globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND).
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 153 AA; 17178 MW; 8DC3BD451EDAA1E9 CRC64;

Query Match 90.9%; Score 722; DB 1; Length 153;
Best Local Similarity 90.1%; Pred. No. 1.8e-55;
Matches 137; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHWAKVEADVAGHGQDILIRLPKSHPTLEKFDPRKHLKTEAEKASDEL 61
Db 2 LSEAEWQLVHWAKVEADLSHGQQLIRLPKSHPTLEKFDPRKHLKTEAEKASDEL 61
QY 62 KKHGTVTLTALGAILKKKGHEAEKPLAQSHATKHPIKYLEFTISEAILHVLHSRHPG 121
Db 62 KKHGTVTLTALGAILKKKGHEAEKPLAQSHATKHPIKYLEFTISEAILHVLHSRHPG 121
QY 122 DFGADAQAMNKALELFRKDIAAKYKELGYQG 153
Db 122 DFGADAQAMNKALELFRKDIAAKYKELGFGH 153

RESULT 15
ID MYG_HORSE STANDARD; PRT; 153 AA.
AC P02183;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myoglobin.
GN MB.
OS Equus caballus (Horse), and
OS Equus burchelli (Plains zebra) (Equus quagga).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796, 9790;
RN [1]_
RP SEQUENCE.
RX SPECIES=Horse; TISSUE=Heart muscle;
RX MEDLINE=70064092; PubMed=4902609;
RA Dautrevaux M., Boulanger Y., Han K., Biserte G.;
RT "Covalent structure of horse myoglobin.";
RL Eur. J. Biochem. 111:267-277(1969).
RN [2]_

```

```

RP SEQUENCE.
RC SPECIES=Horse; TISSUE=Skeletal muscle;
RA Romero-Herrera A.E., Lehmann H.;
RT "Residue 122 of sperm whale and horse myoglobin.";
RL Biochim. Biophys. Acta 336:318-323(1974).
RN [3]_
RP SEQUENCE OF 1-15 AND 56-70.
RC SPECIES=Horse;
RX MEDLINE=90147691; PubMed=2302197;
RA Jahnen W., Ward L.D., Reid G.E., Moritz R.L., Simpson R.J.;
RT "Internal amino acid sequencing of proteins by in situ cyanogen
bromide cleavage in polyacrylamide gels.";
RL Biochem. Biophys. Res. Commun. 166:139-145(1990).
RN [4]_
RP COMPOSITION OF TRYPTIC AND PEPTIC PEPTIDES.
RC SPECIES=E.burchelli;
RX MEDLINE=75184110; PubMed=1095063;
RA Darbre P.D., Romero-Herrera A.E., Lehmann H.;
RT "Comparison of the myoglobin of the zebra (Equus burchelli) with that
of the horse (Equus caballus).";
RL Biochim. Biophys. Acta 393:201-204(1975).
RN [5]_
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RC SPECIES=Horse;
RX MEDLINE=90294302; PubMed=2359126;
RA Evans S.V., Brayer G.D.;
RT "High-resolution study of the three-dimensional structure of horse
heart metmyoglobin.";
RL J. Mol. Biol. 213:885-897(1990).
RN [6]_
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RC SPECIES=Horse;
RX MEDLINE=88153746; PubMed=3346247;
RA Evans S.V., Brayer G.D.;
RT "Horse heart metmyoglobin. A 2.8-A resolution three-dimensional
structure determination.";
RL J. Biol. Chem. 263:4263-4268(1988).
RN [7]_
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF MUTANT THR-64.
RC SPECIES=Horse;
RX MEDLINE=95383313; PubMed=7654702;
RA Boquil R., Maurus R., Hildebrand D.P., Brayer G.D., Mauk A.G.;
RT "Origin of the pH-dependent spectroscopic properties of
pentacoordinate metmyoglobin variants.";
RL Biochemistry 34:10483-10490(1995).
RN [8]_
RP X-RAY CRYSTALLOGRAPHY (1.45 ANGSTROMS).
RC SPECIES=Horse;
RX MEDLINE=20168811; PubMed=10706294;
RA Chu K., Voitchovsky J., McMahon B.H., Sweet R.M., Berendzen J.,
RA Schlichting I.;
RT "Structure of a ligand-binding intermediate in wild-type carbonmonoxy
myoglobin.";
RL Nature 403:921-923(2000).
CC -!- FUNCTION: Serves as a reserve supply of oxygen and facilitates the
movement of oxygen within muscles.
CC -!- SIMILARITY: Belongs to the globin family.
DR PIR; A30603; MYH0Z.
DR PIR; A91098; MYH0.
DR PDB; 1YMA; 31-JAN-94.
DR PDB; 1YMB; 31-JAN-94.
DR PDB; 1YMC; 31-JAN-94.
DR PDB; 1HRM; 26-JAN-95.
DR PDB; 1HSY; 27-FEB-95.
DR PDB; 1RSE; 23-DEC-96.
DR PDB; 1XCH; 17-SEP-97.
DR PDB; 1WLA; 14-JAN-98.
DR PDB; 1BJE; 28-JAN-98.
DR PDB; 1AZI; 25-FEB-98.
DR PDB; 1DWR; 03-MAR-00.
DR PDB; 1DWS; 24-JUN-03.
DR PDB; 1DWT; 03-MAR-00.
DR PDB; 1GJN; 01-MAR-02.

```

```
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle; 3D-structure.
FT METAL 64 64
FT METAL 93 93
FT CONFLICT 122 122
FT HELIX 4 18
FT TURN 19 20
FT HELIX 21 35
FT HELIX 37 40
FT TURN 41 42
FT TURN 44 48
FT HELIX 52 57
FT HELIX 59 77
FT TURN 78 80
FT HELIX 83 94
FT TURN 95 96
FT HELIX 101 118
FT TURN 120 121
FT HELIX 125 148
FT TURN 149 150
SQ SEQUENCE 153 AA; 16951 MW; 89CA01974231E93C CRC64;

Query Match 90.3%; Score 717; DB 1; Length 153;
Best Local Similarity 88.2%; Pred. No. 5e-55;
Matches 134; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

QY 2 LSEGEWQLVHWAKVEADVAGHGQDILIRLFKSHPETLEKFDREKHLKTEAEKASDGL 61
Db 2 LSDGEWQVLNVWGVKEADVAGHGQDILIRLFKSHPETLEKFDREKHLKTEAEKASDGL 61

QY 62 KKHGTVTLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPG 121
Db 62 KKHGTVTLTALGAILKKKGHHEAEKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPG 121

QY 122 DFGADAQGAMNKALELFRKDIAAKYKELGYQG 153
Db 122 DFGADAQGAMTKALELFRNDIAAKYKELGFQG 153
```

Search completed: August 10, 2004, 15:26:50
Job time : 5.81598 secs

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:22:05 ; Search time 18.3378 Seconds
(without alignments)
2632.505 Million cell updates/sec

Title: US-09-455-978b-76
Perfect score: 794
Sequence: 1 VLSEGEWQLVHLHWAKVEAD.....ALELFRKDIAKYKELGYQG 153

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteria:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	647	81.5	154	11	Q9QZ76	Q9qz76 rattus norv
2	460	57.9	99	4	Q8WVH6	Q8wvh6 homo sapien
3	321.5	40.5	147	13	Q9DGI1	Q9dgi1 makaira nig
4	312.5	39.4	147	13	Q9DGI3	Q9dgi3 sarda chile
5	305.5	38.5	147	13	Q9DGI7	Q9dgi7 thunnus alb
6	288.5	36.3	147	13	Q9DGI2	Q9dgi2 thunnus ala
7	288	36.3	146	13	Q9DGI8	Q9dgi8 katsuwonus
8	285.5	36.0	142	13	Q9DWH9	Q9dwh9 hemitriptery
9	283.5	35.7	147	13	Q7T044	Q7t044 channichthy
10	271.5	34.2	141	13	P79846	P79846 pagetopsis
11	267.5	33.7	147	13	Q9DGI9	Q9dgi9 scomber jap
12	237.5	29.9	110	13	Q98963	Q98963 notochenia
13	215.5	27.1	190	11	Q8BN80	Q8bn80 mus musculu
14	211.5	26.6	190	4	Q8N2X5	Q8n2x5 homo sapien
15	194.5	24.5	103	13	Q8AY74	Q8ay74 champsoceph
16	192.5	24.2	103	13	O57410	O57410 champsoceph

17	169.5	21.3	177	13	Q802S6	Q802s6 oncorhynchu
18	146	18.4	147	6	Q9BEI1	Q9beil sminthopsis
19	144.5	18.2	143	13	Q7ZT21	Q7zt21 brachydanio
20	143.5	18.1	143	13	Q7SZV8	Q7szv8 brachydanio
21	143	18.0	142	13	Q9PVL9	Q9pvl9 hyobius re
22	143	18.0	147	6	Q9BEI2	Q9bei2 macropus eu
23	142.5	17.9	143	13	Q8UVI8	Q8uv18 sphoeroides
24	142	17.9	147	11	O88752	O88752 rattus norv
25	140	17.6	144	13	Q8UW52	Q8uw52 cyprinus ca
26	138.5	17.4	147	11	Q9D0B2	Q9d0b2 mus musculu
27	138	17.4	146	11	Q9QUN8	Q9qun8 mus musculu
28	138	17.4	147	11	O9CR49	O9cr49 mus musculu
29	137	17.3	35	6	Q865L4	Q865l4 bos taurus
30	136.5	17.2	143	13	Q8UW95	Q8uw95 cyprinus ca
31	135.5	17.1	143	13	Q8UW95	Q8uw95 cyprinus ca
32	135.5	17.1	143	13	Q8AYQ1	Q8ayq1 oryzias lat
33	134.5	16.9	143	13	Q7SZV9	Q7szv9 brachydanio
34	134	16.9	142	6	Q95233	Q95233 perodicticu
35	134	16.9	142	11	O3CWS5	O3cws5 mus musculu
36	133.5	16.8	143	13	Q803Z5	Q803z5 brachydanio
37	133	16.8	142	11	O91V15	O91v15 rattus norv
38	133	16.8	147	6	Q9GLX5	Q9glx5 aotus nancy
39	132	16.6	142	6	Q28743	Q28743 ovis aries
40	132	16.6	142	13	Q90ZAS	Q90zas ambystoma m
41	132	16.6	147	6	Q9GJS7	Q9gjs7 callithrix
42	132	16.6	147	6	Q9GLX7	Q9glx7 saimiri sci
43	132	16.6	147	13	Q90938	Q90938 gallus gall
44	131.5	16.6	143	13	Q8JH83	Q8jh83 misgurnus a
45	131.5	16.6	143	13	Q8JH82	Q8jh82 paramisgurn

ALIGNMENTS

RESULT 1

Q9QZ76 PRELIMINARY; PRT; 154 AA.
ID Q9QZ76
AC Q9QZ76;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Myoglobin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar;
RA Santos R.A., Giannocco G., Poyares L.L., Nunes M.T.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF197916; AAF05848.1; -
DR HSP; P02189; IMWC.
DR GO; GO:0005344; F: oxygen transporter activity; IEA.
DR GO; GO:0015671; P: oxygen transport; IEA.
DR GO; GO:0006810; P: transport; IEA.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
SQ SEQUENCE 154 AA; 17157 MW; 98B31BBE57FA020DA CRC64;

Query Match 81.5%; Score 647; DB 11; Length 154;
Best Local Similarity 78.9%; Pred. No. 2.1e-52;
Matches 120; Conservative 18; Mismatches 14; Indels 0; Gaps 0;
Qy 2 LSEGEWQLVHLHWAKVEADVAGHGDQILIRLFKSHPETLEKFRFKHLKTAEMKASD 61
Db 3 LSDGEWQVLNIGKVGDLAGHGOEVLISLFKAHPETLEKFRFKHLKSEEMKSSDL 62
Qy 62 KKHGVTVLTALGAILKKKGHEALPLAQSHATKHKPIKYLEFISEAILHVLHSLRHPG 121

```
Db 63 KHGGCTVLTGTLKXGQRAEIQPLAQSHATKHKIPVKYLEFISEVILQVLKRYSG 122
QY 122 DFGADAQAMNKALELFRKDIAAKYKELGYQG 153
Db 123 DFGADAQAMSKALELFRNDIAAKYKELGFGQ 154

RESULT 2
Q8WH6 PRELIMINARY; PRT; 99 AA.
AC Q8WH6;
DT 01-WAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Similar to myoglobin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RA Strauberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC018001; RAH18001.1; -.
DR HSP; P02185; 111M.
DR GO; GO:0005344; P:oxygen transporter activity; IEA.
DR GO; GO:0015671; P:oxygen transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; Globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
SQ SEQUENCE 99 AA; 10870 MW; 73E67F581F79C479 CRC64;

Query Match 57.9%; Score 460; DB 4; Length 99;
Best Local Similarity 87.9%; Pred. No. 3e-35;
Matches 87; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 55 MKASDLKKHGVTVLTALGAILKKGHHEAEIKPLAQSHATKHKIPVKYLEFISEBATHV 114
Db 1 MKASDLKKHGAIVTLTGILKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECTIOV 60

QY 115 LHSRHPGDFGADAQAMNKALELFRKDIAAKYKELGYQG 153
Db 61 LQSKHPGDFGADAQAMNKALELFRKDMASNYKELGFGQ 99

RESULT 3
Q9DGJ1 PRELIMINARY; PRT; 147 AA.
AC Q9DGJ1;
DT 01-WAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Myoglobin.
OS Makaira nigricans (Blue marlin).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes; Scombroidei;
OC Xiphidae; Makaira.
OX NCBI_TaxID=13604;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=21146139; PubMed=11247835;
RA Marcinek D.J., Bonaventura J., Wittenberg J.B., Block B.A.;
RT "Oxygen affinity and amino acid sequence of myoglobins from endothermic and ectothermic fish.";
RL Am. J. Physiol. Regul. Integr. Comp. Physiol. 280:R1123-R1133(2001).

Query Match 39.4%; Score 312.5; DB 13; Length 147;
Best Local Similarity 44.6%; Pred. No. 2.5e-21;
Matches 66; Conservative 21; Mismatches 58; Indels 3; Gaps 2;

QY 6 EQWLVLHVWAKVEADVAGHGDILIRLPKSHPETLEKFDPRFKHLKTEAMKASEDLKKHG 65
Db 3 DFDVILKFWGPVEADYTHGNVLTRLTETHPETQKLPKFAGI-AKADMAGNAAISAHG 61
```

```
DR EMBL; AF291833; AAG02107.1; -.
DR HSP; P02205; 1MYT.
DR GO; GO:0005344; P:oxygen transporter activity; IEA.
DR GO; GO:0015671; P:oxygen transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
SQ SEQUENCE 147 AA; 15842 MW; F52D010973F4D84B CRC64;

Query Match 40.5%; Score 321.5; DB 13; Length 147;
Best Local Similarity 46.7%; Pred. No. 3.7e-22;
Matches 70; Conservative 20; Mismatches 53; Indels 7; Gaps 3;

QY 6 EQWLVLHVWAKVEADVAGHGDILIRLPKSHPETLEKFDPRFKHLKTEAMKASEDLKKHG 65
Db 3 DFEVLKHWGPVEADYATHGNVLTRLTETHPETQKLPKFAGI-AKADMAGNAAISAHG 61

QY 66 VTVLTAALGAILKKGHHEAEIKPLAQSHATKHKIPVKYLEFISEBATHVILHSRHPGDFGA 125
Db 62 ATVLKKGELKAKGSHAAIKPMANSHATKHKIPKPFELISEVIGVMHEK---AGL 117

QY 126 DAQG--AMNKALELFRKDIAAKYKELGYQG 153
Db 118 DAQGQALKNVMTTIIADIEANYKELGFTG 147

RESULT 4
Q9DGJ0 PRELIMINARY; PRT; 147 AA.
AC Q9DGJ0;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Myoglobin.
OS Sarda chiliensis (Sard).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes; Scombroidei;
OC Scombridae; Sarda.
OX NCBI_TaxID=8231;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=21146139; PubMed=11247835;
RA Marcinek D.J., Bonaventura J., Wittenberg J.B., Block B.A.;
RT "Oxygen affinity and amino acid sequence of myoglobins from endothermic and ectothermic fish.";
RL Am. J. Physiol. Regul. Integr. Comp. Physiol. 280:R1123-R1133(2001).
DR EMBL; AF291834; AAG02108.1; -.
DR HSP; P02205; 1MYT.
DR GO; GO:0005344; P:oxygen transporter activity; IEA.
DR GO; GO:0015671; P:oxygen transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
SQ SEQUENCE 147 AA; 15797 MW; D0864510E730506 CRC64;

Query Match 39.4%; Score 312.5; DB 13; Length 147;
Best Local Similarity 44.6%; Pred. No. 2.5e-21;
Matches 66; Conservative 21; Mismatches 58; Indels 3; Gaps 2;

QY 6 EQWLVLHVWAKVEADVAGHGDILIRLPKSHPETLEKFDPRFKHLKTEAMKASEDLKKHG 65
Db 3 DFDVILKFWGPVEADYTHGNVLTRLTETHPETQKLPKFAGI-AQADMAGNAAISAHG 61
```



```

QY 66 VTVLTALGAILKKKGHEAEALPKLAOSHATKHKIPKYLEFISEAIHVLHSHRHPDGA 125
Db 62 ATVLKGLGELLKAGNHAAILKPMANSHATKHKIPINNFKLISEIIVKVMQEKAGDAG- 120

QY 126 DAQGANMKALELFRKIDIAKYKELGYQG 153
Db 121 -GQALRNVMGIIADLEANYKELGFSG 147

RESULT 5
Q9DGI7 PRELIMINARY; PRT; 147 AA.
AC Q9DGI7;
RC TISSUE=Skeletal muscle;
RX MEDLINE=21146139; PubMed=11247835;
RA Marcinek D.J., Bonaventura J., Wittenberg J.B., Block B.A.;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE Myoglobin.
OS Thunnus albacares (Yellowfin tuna) (Neothunnus macropterus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;
OC Scombridae; Thunnus.
OX NCBI_TaxID=8236;
RN [1]
RP SEQUENCE FROM N.A.
QY 66 VTVLTALGAILKKKGHEAEALPKLAOSHATKHKIPKYLEFISEAIHVLHSHRHPDGA 125
Db 62 ATVLKGLGELLKAGNHAAILKPMANSHATKHKIPINNFKLISEIIVKVMQEKAGDAG- 120

Query Match 38.5%; Score 305.5; DB 13; Length 147;
Best Local Similarity 43.9%; Pred. No. 1.1e-20;
Matches 65; Conservative 22; Mismatches 58; Indels 3; Gaps 2;

QY 6 EWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEKASEDLKXHG 65
Db 3 DFDAVLKCGPVEADYTTMGVLTLRFLKEHPTQKLPFKFAGI-AQADIAGNAAISAHG 61

QY 66 VTVLTALGAILKKKGHEAEALPKLAOSHATKHKIPKYLEFISEAIHVLHSHRHPDGA 125
Db 62 ATVLKGLGELLKAGSHAAILKPLANSATKHKIPINNFKLISEIIVKVMQEKAGDAG- 120

QY 126 DAQGANMKALELFRKIDIAKYKELGYQG 153
Db 121 -GQALRNVMGIIADLEANYKELGFSG 147

Query Match 36.3%; Score 288.5; DB 13; Length 147;
Best Local Similarity 41.2%; Pred. No. 4.4e-19;
Matches 61; Conservative 25; Mismatches 59; Indels 3; Gaps 2;

QY 6 EWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEKASEDLKXHG 65
Db 3 DFDAVLKCGPVEADYTTMGVLTLRFLKEHPTQKLPFKFAGI-AQADIAGNAAISAHG 61

QY 66 VTVLTALGAILKKKGHEAEALPKLAOSHATKHKIPKYLEFISEAIHVLHSHRHPDGA 125
Db 62 ATVLKGLGELLKAGSHAAILKPLANSATKHKIPINNFKLISEIIVKVMQEKAGDAG- 120

QY 126 DAQGANMKALELFRKIDIAKYKELGYQG 153
Db 121 -GQALRNVMGIIADLEANYKELGFTG 147

RESULT 7
Q9DGI8 PRELIMINARY; PRT; 146 AA.
AC Q9DGI8;
RC TISSUE=Skeletal muscle;
RX MEDLINE=21146139; PubMed=11247835;
RA Marcinek D.J., Bonaventura J., Wittenberg J.B., Block B.A.;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE Myoglobin.
OS Katsuwonus pelamis (Skipjack tuna) (Bonito).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;
OC Scombridae; Katsuwonus.
OX NCBI_TaxID=8226;
RN [1]
RP SEQUENCE FROM N.A.
QY 66 VTVLTALGAILKKKGHEAEALPKLAOSHATKHKIPKYLEFISEAIHVLHSHRHPDGA 125
Db 62 ATVLKGLGELLKAGSHAAILKPLANSATKHKIPINNFKLISEIIVKVMQEKAGDAG- 120

QY 126 DAQGANMKALELFRKIDIAKYKELGYQG 153
Db 121 -GQALRNVMGIIADLEANYKELGFTG 147

RESULT 6
Q9DGI2 PRELIMINARY; PRT; 147 AA.
AC Q9DGI2;
RC TISSUE=Skeletal muscle;
RX MEDLINE=21146139; PubMed=11247835;
RA Marcinek D.J., Bonaventura J., Wittenberg J.B., Block B.A.;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE Myoglobin.
OS Thunnus alalunga (Albacore).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

```

```

OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;
OC Scombridae; Thunnus.
OX NCBI_TaxID=8235;
RN [1]
RP SEQUENCE FROM N.A.
QY 66 VTVLTALGAILKKKGHEAEALPKLAOSHATKHKIPKYLEFISEAIHVLHSHRHPDGA 125
Db 62 ATVLKGLGELLKAGNHAAILKPMANSHATKHKIPINNFKLISEIIVKVMQEKAGDAG- 120

QY 126 DAQGANMKALELFRKIDIAKYKELGYQG 153
Db 121 -GQALRNVMGIIADLEANYKELGFSG 147

Query Match 36.3%; Score 288.5; DB 13; Length 147;
Best Local Similarity 41.2%; Pred. No. 4.4e-19;
Matches 61; Conservative 25; Mismatches 59; Indels 3; Gaps 2;

QY 6 EWQLVHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRFXHLKTEAEKASEDLKXHG 65
Db 3 DFDAVLKCGPVEADYTTMGVLTLRFLKEHPTQKLPFKFAGI-AQADIAGNAAISAHG 61

QY 66 VTVLTALGAILKKKGHEAEALPKLAOSHATKHKIPKYLEFISEAIHVLHSHRHPDGA 125
Db 62 ATVLKGLGELLKAGSHAAILKPLANSATKHKIPINNFKLISEIIVKVMQEKAGDAG- 120

QY 126 DAQGANMKALELFRKIDIAKYKELGYQG 153
Db 121 -GQALRNVMGIIADLEANYKELGFTG 147

RESULT 7
Q9DGI8 PRELIMINARY; PRT; 146 AA.
AC Q9DGI8;
RC TISSUE=Skeletal muscle;
RX MEDLINE=21146139; PubMed=11247835;
RA Marcinek D.J., Bonaventura J., Wittenberg J.B., Block B.A.;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE Myoglobin.
OS Katsuwonus pelamis (Skipjack tuna) (Bonito).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;
OC Scombridae; Katsuwonus.
OX NCBI_TaxID=8226;
RN [1]
RP SEQUENCE FROM N.A.
QY 66 VTVLTALGAILKKKGHEAEALPKLAOSHATKHKIPKYLEFISEAIHVLHSHRHPDGA 125
Db 62 ATVLKGLGELLKAGSHAAILKPLANSATKHKIPINNFKLISEIIVKVMQEKAGDAG- 120

QY 126 DAQGANMKALELFRKIDIAKYKELGYQG 153
Db 121 -GQALRNVMGIIADLEANYKELGFTG 147

RESULT 6
Q9DGI2 PRELIMINARY; PRT; 147 AA.
AC Q9DGI2;
RC TISSUE=Skeletal muscle;
RX MEDLINE=21146139; PubMed=11247835;
RA Marcinek D.J., Bonaventura J., Wittenberg J.B., Block B.A.;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE Myoglobin.
OS Thunnus alalunga (Albacore).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

```

KW Heme; Oxygen transport; Transport.
SQ SEQUENCE 146 AA; 15396 MW; 307DAA2C6D9FDC27 CRC64;

Query Match 36.3%; Score 288; DB 13; Length 146;
Best Local Similarity 43.8%; Pred. No. 4.8e-19;
Matches 64; Conservative 21; Mismatches 53; Indels 8; Gaps 3

QY 10 VLHWAKVEADVAGHGQDILIRLFKSPHETLEKFDPRFKHLKTEAMKASEDLKXHGVTVL 69
Db 7 VLKCGAVEADVNTVGGVLARLFKDPETQKLFKPKFAGI--TGDAGNAAVAAGATVL 64

QY 70 TALGAILKKGGHBAELKPLAQSHATKHKIPDKYLEFISEALIHVLHSRHPGDFGADAQG 129
Db 65 XKGELKAKGNHAAIITKPLANSHAKQHKIPINNFKLITEALAHVLHEK---AGLDAAG 120

QY 130 --AMNKALELFRKDIAAKYKELGYQG 153
Db 121 QTLRNVNGIIVADLEANYKELGFTG 146

RESULT 8
Q90WH9 PRELIMINARY; PRT; 142 AA.

ID Q90WH9 AC Q90WH9; TISSUE=Heart ventricle;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Myoglobin (Fragment)
OS Hemitripterus americanus (Sea raven).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Scorpaeniformes;
OC Cottoidei; Hemitripterae; Hemitripterus.
OX NCBI_TaxID=8094;
RN [1]

SEQUENCE FROM N.A.
RC TISSUE=Heart ventricle;
RA Grove T.J., Sidell B.D.;
RT "Myoglobin from sea raven, Hemitripterus americanus."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY029587; AAK49781.1; --
DR HSSP; P2185; 111M.
DR GO; GO:0005344; F: oxygen transporter activity; IEA.
DR GO; GO:0015671; P: oxygen transport; IEA.
DR GO; GO:0006810; P: transport; IEA.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
FT NON TER 1
SQ SEQUENCE 142 AA; 15111 MW; 2DB1EF4602F929D3 CRC64;

Query Match 36.0%; Score 285.5; DB 13; Length 142;
Best Local Similarity 44.2%; Pred. No. 7.9e-19;
Matches 65; Conservative 18; Mismatches 57; Indels 7; Gaps 3

QY 9 LVLHWAKVEADVAGHGQDILIRLFKSPHETLEKFDPRFKHLKTEAMKASEDLKXHGVTV 68
Db 1 LVLKCGVPEADYAAAYGSLVLTFRTEHPDQKLFKPKFAGI--AQGDMAADAGISAGGATV 59

QY 69 LTALGAILKKGGHBAELKPLAQSHATKHKIPDKYLEFISEALIHVLHSRHPGDFGADAQG 128
Db 60 LRKUGELKAKGNHAAIITKPLANSHATKHKIPINNFKLITEALAHVLHEK---TGLDAA 115

QY 129 G--AMNKALELFRKDIAAKYKELGYQG 153
Db 116 GQALRNWMAIVVADLEANYKELGFTG 142

RESULT 9
77T044

```

Query Match      34.2%; Score 271.5; DB 13; Length 141;
Best Local Similarity 42.3%; Pred. No. 1.6e-17;
Matches 60; Conservative 20; Mismatches 59; Indels 3; Gaps 2;

QY 10 VLHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRPHLKTAEAMKASEDLKKHGVTVL 69
Db 1 VLKWCMPMEADYATHGGLVTLRFTFHPETLKLFPKFAGI-AHGDLAGDAGVSAHGATVL 59
QY 70 TALGAILKKKGHEAEALPLAQSHATKHKIPKYLEFISEAIIHVLHSHRPGDAG 129
Db 60 NKGDLKARGAHAALLKPLSSSHATKHKIPINFKLIAEVIGKVMEEKAGVDAG--GQT 117

QY 130 ANKALELFRKDIKAAKYLEGY 151
Db 118 GLRNVMAVIIADMEAYKELGF 139

RESULT 11
ID Q9DGI9 PRELIMINARY; PRT; 147 AA.
AC Q9DGI9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Myoglobin.
OS Scomber japonicus (Chub mackerel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;
OC Scombridae; Scomber.
OX NCBI_TaxID=13676;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=21146139; PubMed=11247835;
RA Marcinek D.J., Bonaventura J., Wittenberg J.B., Block B.A.;
RT "Oxygen affinity and amino acid sequence of myoglobins from
endothermic and ectothermic fish.";
RL Am. J. Physiol. Regul. Integr. Comp. Physiol. 280:R1123-R1133 (2001).
DR EMBL: AF291835; AAG02109.1; -.
DR HSSP: P02205; IMYT.
DR GO: GO:0005344; P:Oxygen transporter activity; IEA.
DR GO: GO:0015671; P:Oxygen transport; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR000971; Globin.
DR InterPro: IPR002335; Myoglobin.
DR Pfam: PF00042; Globin; 1.
DR PRINTS: PR00613; MYOGLOBIN.
DR PROSITE: PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
SQ SEQUENCE 147 AA; 15766 MW; 44EB9A4611EE0366 CRC64;

Query Match      33.7%; Score 267.5; DB 13; Length 147;
Best Local Similarity 39.9%; Pred. No. 3.9e-17;
Matches 59; Conservative 23; Mismatches 63; Indels 3; Gaps 2;

QY 6 EWOLVHWAKVEADVAGHGQDILIRLFKSHPETLEKDFRPHLKTAEAMKASEDLKKHG 65
Db 3 DFDVAVLKWFGPVEADYDKIGNVLTRLTFTFHPETLKLFPKFAGI-GDMAGNAISAHG 61
QY 66 VTVLTALGAILKKKGHEAEALPLAQSHATKHKIPKYLEFISEAIIHVLHSHRPGDAG 125
Db 62 ATVLKKAELVAKAGNHAGIIRPLANSATKHKIIANNFKLITELIVKMQEKAGLDAG- 120

QY 126 DAQGANWKALELFRKDIKAAKYLEGYQ 153
Db 121 -GQTALRNVMGVFIADMDANYKELGFSG 147

RESULT 12
Q98963
ID Q98963 PRELIMINARY; PRT; 110 AA.
AC Q98963;

```

```

DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Myoglobin (fragment).
OS Notothenia coriiceps (black rockcod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
OC Nototheniidae; Nototheniidae; Notothenia.
OX NCBI_TaxID=8208;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart ventricle;
RA Vayda M.E., Small D.J., Sidell B.D.;
RT "Expression of the myoglobin gene in Antarctic channichthyid fishes.";
RL (in) Battaglia B., Valencia J., Walton S.W.H. (eds.);
VI SCAR ANTARCTIC COMMUNITIES, pp.1-1, Cambridge University Press,
Cambridge, UK (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart ventricle;
RX MEDLINE=98360008; PubMed=9694664;
RA Small D.J., Vayda M.E., Sidell B.D.;
RT "A novel vertebrate myoglobin gene containing three A+T-rich introns
is conserved among Antarctic teleost species which differ in myoglobin
expression.";
RL J. Mol. Evol. 47:156-166 (1998).
DR EMBL: U68350; AAC69245.1; -.
DR HSSP: P02205; IMYT.
DR GO: GO:0005344; P:Oxygen transporter activity; IEA.
DR GO: GO:0015671; P:Oxygen transport; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR000971; Globin.
DR Pfam: PF00042; Globin; 1.
DR PROSITE: PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport.
FT NON_TER 1
FT NON_TER 110
SQ SEQUENCE 110 AA; 11630 MW; B43B5865F0E6708F CRC64;

Query Match      29.9%; Score 237.5; DB 13; Length 110;
Best Local Similarity 45.9%; Pred. No. 1.7e-14;
Matches 50; Conservative 15; Mismatches 43; Indels 1; Gaps 1;

QY 10 VLHVWAKVEADVAGHGQDILIRLFKSHPETLEKDFRPHLKTAEAMKASEDLKKHGVTVL 69
Db 1 VLKWCMPMEADYATHGGLVTLRFTFHPETLKLFPKFAGI-AHGDLAGDAGVSAHGATVL 59
QY 70 TALGAILKKKGHEAEALPLAQSHATKHKIPKYLEFISEAIIHVLHSHR 118
Db 60 NKGDLKARGAHAALLKPLSSSHATKHKIPINFKLIAEVIGKVMEEK 108

RESULT 13
Q98N80
ID Q98N80 PRELIMINARY; PRT; 190 AA.
AC Q98N80;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Mus musculus cytoglobin.
GN CYGB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Eye;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:18:30 ; Search time 31.8547 Seconds
(without alignments)
1632.058 Million cell updates/sec

Title: US-09-455-978b-77
Perfect score: 933
Sequence: 1 MSNDNDTLVTADVRNGIDGH.....DELVARFLPMLKLTFFDQI 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	933	100.0	489	4	AAE04641	Aae04641 Halobacte
2	203	21.8	39	4	AAE04677	Aae04677 Haem prot
3	142	15.2	432	4	AAE04642	Aae04642 Bacillus
4	99	10.6	448	1	ABP70493	Abp70493 Protein G
5	99	10.6	448	1	ABP95030	Abp95030 Protein G
6	99	10.6	448	2	AAE07013	Aae07013 Protein G
7	99	10.6	448	2	AAE10004	Aae10004 Streptoco
8	99	10.6	448	2	AAE53290	Aae53290 Streptoco
9	99	10.6	1175	2	AAE34570	Aae34570 Porphorym
10	99	10.6	1232	2	AAE34569	Aae34569 Porphorym
11	99	10.6	1266	2	AAE34568	Aae34568 Porphorym
12	99	10.6	1269	2	AAE34428	Aae34428 Porphorym
13	98	10.5	593	2	AAE07014	Aae07014 Protein G
14	98	10.5	593	2	AAE62944	Aae62944 Streptoco
15	98	10.5	594	2	AAE10005	Aae10005 Streptoco
16	89	9.5	955	2	AAE57365	Aae57365 K39 polyp
17	89	9.5	955	2	AAE03691	Aae03691 Leishmani
18	88	9.4	505	6	ABU38101	Abu38101 Protein e
19	87	9.3	496	6	ABU37440	Abu37440 Protein e
20	87	9.3	505	6	ABP79309	Abp79309 N. gonorr
21	87	9.3	523	3	AAE34454	Aae34454 Porphorym
22	87	9.3	533	2	AAE34330	Aae34330 Porphorym
23	86.5	9.3	318	3	AAE07677	Aae07677 Amino aci
24	86	9.2	483	7	ADC31377	Adc31377 Human nov
25	85.5	9.2	302	6	ADA33559	Ada33559 Acinetoba

RESULT 1

AAE04641	ID	AAE04641	standard; protein; 489 AA.
XX	AC	AAE04641;	
XX	DT	11-SEP-2003 (revised)	
XX	DT	04-SEP-2001 (first entry)	
XX	DE	Halobacterium salinarum HemAT-Hs protein.	
XX	KW	Haem binding protein; HemAT-Hs; HemAT-Bs; gaseous ligand sensor;	
XX	KW	oxygen storage; artificial photosynthesis; signalling function;	
XX	KW	alpha-haemoglobin; myoglobin; therapy.	
XX	OS	Halobacterium salinarum.	
XX	PN	WO200140475-A2.	
XX	PD	07-JUN-2001.	
XX	PF	05-DEC-2000; 2000WO-US033048.	
XX	PR	06-DEC-1999; 99US-00455978.	
XX	PA	(UYHA-) UNIV HAWAII.	
XX	PI	Alam M, Larsen R;	
XX	DR	WPI; 2001-374832/39.	
XX	DR	N-PSDB; AAD08991.	
XX	PT	Novel isolated bacterial heme binding protein, HemAT-Hs or HemAT-Bs which reversibly binds oxygen with low affinity, useful for controlled storage of oxygen and for sensing gaseous ligands such as oxygen.	
XX	PS	Claim 6; Page 10; 94pp; English.	
XX	CC	The present invention relates to isolated archeal and bacterial haem binding protein, HemAT-Hs or HemAT-Bs which reversibly binds oxygen with low affinity. Haem binding protein is useful for controlled storage of oxygen by allowing haem binding protein to bind and store oxygen, and triggering the release of oxygen from haem binding protein by activating the signalling domain. Haem binding protein is useful for sensing gaseous ligands such as oxygen, NO, CO, or CN. Blood substitute comprising haem binding protein is useful for treating a patient suffering from low blood levels by administering and regulating the oxygen binding of the haem-binding protein by modifying the signalling domain. Haem binding protein	

CC is useful for haem-based catalysis, for artificial photosynthesis and for
CC identifying potential signalling functions of mutated alpha-haemoglobin
CC and myoglobin causing several diseases. The present sequence is
CC Halobacterium salinarium HemAT-Hs protein which is salt tolerant.
CC (Updated on 11-SEP-2003 to standardise OS field)

XX Sequence 489 AA;

Query Match 100.0%; Score 933; DB 4; Length 489;
Best Local Similarity 100.0%; Pred. No. 2.6e-89;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNDNTLVADVRGIDGHALADRGIDAEIAHRLSFTGIDDDTMAALAAEQPLFEAT 60
|||||
DB 1 MSNDNTLVADVRGIDGHALADRGIDAEIAHRLSFTGIDDDTMAALAAEQPLFEAT 60
QY 61 ADALVTDFDHLSEYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIG 120
|||||
DB 61 ADALVTDFDHLSEYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIG 120
QY 121 KTHDVLGLGPDVYLGAYTRYTGLLDALADDDVADRGEEAAAADDELVARFLPMLKLTFF 180
|||||
DB 121 KTHDVLGLGPDVYLGAYTRYTGLLDALADDDVADRGEEAAAADDELVARFLPMLKLTFF 180

QY 181 DQOI 184
|||||

DB 181 DQOI 184

RESULT 2

AAE04677
ID AAE04677 standard; peptide; 39 AA.

XX AAE04677;

AC AAE04677;

DT 11-SEP-2003 (revised)

DT 04-SEP-2001 (first entry)

XX Haem protein related Halobacterium salinarium protein fragment.

XX Haem binding protein; HemAT-Hs; HemAT-Bs; gaseous ligand sensor;

KW oxygen storage; artificial photosynthesis; signalling function;

KW alpha-haemoglobin; myoglobin; therapy.

XX Halobacterium salinarum.

XX Key Location/Qualifiers

FF 19..39

FT Domain /label= M2_Box

XX WO200140475-A2.

XX 07-JUN-2001.

XX 05-DEC-2000; 2000WO-US033048.

XX 06-DEC-1999; 99US-00455978.

XX (UYHA-) UNIV HAWAII.

XX Alam M, Larsen R;

XX WPI; 2001-374832/39.

XX Novel isolated bacterial heme binding protein, HemAT-Hs or HemAT-Bs which
PT reversibly binds oxygen with low affinity, useful for controlled storage
PT of oxygen and for sensing gaseous ligands such as oxygen.

PS Disclosure; Page 16; 94pp; English.

XX The present invention relates to isolated archeal and bacterial haem
CC binding protein, HemAT-Hs or HemAT-Bs which reversibly binds oxygen with
CC low affinity. Haem binding protein is useful for controlled storage of

CC oxygen by allowing haem binding protein to bind and store oxygen, and
CC triggering the release of oxygen from haem binding protein by activating
CC the signalling domain. Haem binding protein is useful for sensing gaseous
CC ligands such as oxygen, NO, CO, or CN. Blood substitute comprising haem
CC binding protein is useful for treating a patient suffering from low blood
CC levels by administering and regulating the oxygen binding of the haem-
CC binding protein by modifying the signalling domain. Haem binding protein
CC is useful for haem-based catalysis, for artificial photosynthesis and for
CC identifying potential signalling functions of mutated alpha-haemoglobin
CC and myoglobin causing several diseases. The present sequence is a haem
CC protein related Halobacterium salinarium protein fragment. (Updated on 11
CC -SEP-2003 to standardise OS field)

XX Sequence 39 AA;

Query Match 21.8%; Score 203; DB 4; Length 39;
Best Local Similarity 100.0%; Pred. No. 6e-14;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 QAEYLLGLGRGEYDTEYAAQRARIGKIHDLVGLGPDVYL 134
|||||

DB 1 QAEYLLGLGRGEYDTEYAAQRARIGKIHDLVGLGPDVYL 39
|||||

RESULT 3

AAE04642

ID AAE04642 standard; protein; 432 AA.

XX AAE04642;

XX 04-SEP-2001 (first entry)

XX Bacillus subtilis HemAT-Bs protein.

XX Haem binding protein; HemAT-Hs; HemAT-Bs; gaseous ligand sensor;

KW oxygen storage; artificial photosynthesis; signalling function;

KW alpha-haemoglobin; myoglobin; therapy.

XX Bacillus subtilis.

XX WO200140475-A2.

XX 07-JUN-2001.

XX 05-DEC-2000; 2000WO-US033048.

XX 06-DEC-1999; 99US-00455978.

XX (UYHA-) UNIV HAWAII.

XX Alam M, Larsen R;

XX WPI; 2001-374832/39.

DR N-PSDB; AAD08992.

XX Novel isolated bacterial heme binding protein, HemAT-Hs or HemAT-Bs which
PT reversibly binds oxygen with low affinity, useful for controlled storage
PT of oxygen and for sensing gaseous ligands such as oxygen.

XX Claim 8; Page 11; 94pp; English.

XX The present invention relates to isolated archeal and bacterial haem
CC binding protein, HemAT-Hs or HemAT-Bs which reversibly binds oxygen of
CC low affinity. Haem binding protein is useful for controlled storage of
CC oxygen by allowing haem binding protein to bind and store oxygen, and
CC triggering the release of oxygen from haem binding protein by activating
CC the signalling domain. Haem binding protein is useful for sensing gaseous
CC ligands such as oxygen, NO, CO, or CN. Blood substitute comprising haem
CC binding protein is useful for treating a patient suffering from low blood
CC levels by administering and regulating the oxygen binding of the haem-
CC binding protein by modifying the signalling domain. Haem binding protein
CC is useful for haem-based catalysis, for artificial photosynthesis and for
CC identifying potential signalling functions of mutated alpha-haemoglobin

Query Match
10.6%; Score 99; DB 1; Length 448;


```

XX SQ      Sequence 448 AA;
      Query Match      10.6%; Score 99; DB 2; Length 448;
      Best Local Similarity 25.3%; Pred. No. 0.17;
      Matches 48; Conservative 25; Mismatches 45; Indels 72; Gaps 8;
QY 10 TADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDD-----DTMAALAAE----- 53
Db 41 TPIIRNGGE--LTNLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAAWAEA 97
QY 54 -----QPLFEATADAL-----VTDYDHLSEYERTQDLFANSTKIVTEOLKETQAEYLLG 102
Db 98 AAAADALAKAKADALKEFNKYGVSDYKKNL-----INNAKTVEGIKDLQAQV-- 145
QY 103 LGRGEYDTEYAAQRARIGKIHDLVGLGPDVYLGAIVYTYTGLLDALADDDVADRGEAA 162
Db 146 -----ESAKKARISEATD-----GLSDFLKSQTPA---EDTVK 175
QY 163 AVDELVARFL 172
Db 176 SIELAEAKVL 185

RESULT 8
AAARS3290
ID AARS3290 standard; protein; 448 AA.
AC AARS3290;
XX
DT 16-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 06-JAN-1995 (first entry)
XX
DE Streptococcus Protein G derived from strain GX7809.
XX
KW Streptococcus Protein G; variant; IgG binding activity; immunoglobulin;
KW Lancefield Group G; bacterial Fc receptor.
XX
OS Streptococcus sp. GX7805.
XX
FH Key Location/Qualifiers
FT Active-site 228..282
FT Region 283..297
FT /label= B1
FT /label= b
FT /note= "linking region"
FT Active-site 298..352
FT /label= B2
XX
XX US5312901-A.
XX
XX 17-MAY-1994.
XX
XX 21-APR-1992; 92US-00871539.
XX
XX 14-FEB-1986; 86US-00829354.
XX 23-APR-1986; 86US-00854887.
XX 19-JUN-1987; 87US-00063959.
XX 20-JUN-1988; 88US-00209236.
XX 19-JUN-1990; 90US-00540169.
XX
XX (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.
XX
XX Farnestock SR;
XX
XX WPI; 1994-159179/19.
XX N-PSDB; AAQ64644.
XX
XX New recombinant streptococcal protein G variants - useful for antibody
XX detection and purification and for therapy.
XX
XX Example 2; Fig 3 and Fig 8; 48pp; English.
PS

```

```

XX CC      A 1.9kb HindIII fragment containing the entire coding sequence for
CC Protein G was isolated from Streptococcus GX7809. The Protein G has IgG-
CC binding activity which has been localised to the B repeating structure.
CC Streptococcal Protein G variants comprising the B domains are claimed.
CC (Updated on 25-MAR-2003 to correct PF field.) (Updated on 16-OCT-2003 to
CC standardise OS field)
XX SQ      Sequence 448 AA;
      Query Match      10.6%; Score 99; DB 2; Length 448;
      Best Local Similarity 25.3%; Pred. No. 0.17;
      Matches 48; Conservative 25; Mismatches 45; Indels 72; Gaps 8;
QY 10 TADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDD-----DTMAALAAE----- 53
Db 41 TPIIRNGGE--LTNLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAAWAEA 97
QY 54 -----QPLFEATADAL-----VTDYDHLSEYERTQDLFANSTKIVTEOLKETQAEYLLG 102
Db 98 AAAADALAKAKADALKEFNKYGVSDYKKNL-----INNAKTVEGIKDLQAQV-- 145
QY 103 LGRGEYDTEYAAQRARIGKIHDLVGLGPDVYLGAIVYTYTGLLDALADDDVADRGEAA 162
Db 146 -----ESAKKARISEATD-----GLSDFLKSQTPA---EDTVK 175
QY 163 AVDELVARFL 172
Db 176 SIELAEAKVL 185

RESULT 9
AAAY34570
ID AAY34570 standard; protein; 1175 AA.
XX
AC AAY34570;
XX
DT 27-AUG-2003 (revised)
DT 20-MAR-2003 (revised)
DT 25-AUG-1999 (first entry)
XX
XX Porphyromonas gingivalis protein PG9.
XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis; vaccine;
XX antigenic.
XX Porphyromonas gingivalis.
XX WO9929870-A1.
XX
XX 17-JUN-1999.
XX
XX 10-DEC-1998; 98WO-AU001023.
XX
XX 10-DEC-1997; 97AU-00000839.
XX 31-DEC-1997; 97AU-00001182.
XX 30-JAN-1998; 98AU-00001546.
XX 10-MAR-1998; 98AU-00002264.
XX 09-APR-1998; 98AU-00002911.
XX 23-APR-1998; 98AU-00003128.
XX 05-MAY-1998; 98AU-00003338.
XX 22-MAY-1998; 98AU-00003654.
XX 29-JUL-1998; 98AU-00004917.
XX 30-JUL-1998; 98AU-00004963.
XX 04-AUG-1998; 98AU-00005028.
XX
XX (CSLUC-) CSL LTD.
XX
XX Ross BC, Barr IG, Patterson MA, Agius CT, Rothel LJ, Margetts MB;
XX Hocking DM, Webb EA;
XX WPI; 1999-385613/32.
XX N-PSDB; AAX91788.
XX

```

XX Antigenic Porphyromonas gingivalis peptides for preventing gingivitis.
 PT Claim 1; Page 565-567; 588pp; English.
 PS AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 XX Porphyromonas gingivalis (PG) polypeptide sequences given in AAX34318 to
 CC AAX34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis. (Updated on 20-MAR-2003 to correct PR field.)
 CC (Updated on 27-AUG-2003 to correct OS field.)
 XX Sequence 1175 AA;
 SQ Query Match 10.6%; Score 99; DB 2; Length 1175;
 Best Local Similarity 25.6%; Pred. No. 0.66;
 Matches 46; Conservative 23; Mismatches 47; Indels 64; Gaps 9;
 QY 16 GIDGHALADRI-----GLDEAEIATWRLSFTG-----IDDDTMAALAAEQPL---- 56
 DB 333 GLDGLASLTRLRRNQISKLEGLDRKLVRLKLDVSGNDIQSIDDIKLLAPILEQTLEKL 392
 QY 57 -----FEATADALVTDFYDHLSEYRTQDLFANSTKTVEQLKETQAEY-----LLG- 102
 DB 393 RIHDNPFVASSGILISPDYDNL-----PEIKALLEKEKEKOKTSVEYHPFCVMLLGN 446
 QY 103 --LGR-----GEYDTEYAAQARIGIKHIVLGL-----GPDVYLGAYTRYT 142
 DB 447 HSSGKTFLSQYDNTNYQ-----KNTHVLSIHRSNPNNAIFYDFGGQDYHYGIIQAFT 501
 RESULT 10
 AAY34569
 ID AAY34569 standard; protein; 1232 AA.
 XX AAY34569;
 XX 27-AUG-2003 (revised)
 DT 20-MAR-2003 (revised)
 DT 25-AUG-1999 (first entry)
 XX Porphyromonas gingivalis protein PG9.
 DE Porphyromonas gingivalis
 XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis; vaccine;
 KW antigenic.
 XX Porphyromonas gingivalis.
 OS WO9929870-A1.
 PN 17-JUN-1999.
 PD 10-DEC-1998; 98WO-AU001023.
 XX 10-DEC-1997; 97AU-00000839.
 PR 31-DEC-1997; 97AU-00001182.
 PR 30-JAN-1998; 98AU-00001546.
 PR 10-MAR-1998; 98AU-00002264.
 PR 09-APR-1998; 98AU-00002911.
 PR 23-APR-1998; 98AU-00003128.
 PR 05-MAY-1998; 98AU-00003338.
 PR 22-MAY-1998; 98AU-00003654.
 PR 29-JUL-1998; 98AU-00004917.
 PR 30-JUL-1998; 98AU-00004963.
 PR 04-AUG-1998; 98AU-00005028.
 XX (CSLC-) CSL LTD.
 XX Ross BC, Barr IG, Patterson MA, Agius CT, Rothel LJ, Margetts MB;
 PI

PI Hocking DM, Webb EA;
 XX WPI: 1999-385613/32.
 DR N-PSDB; AAX91787.
 XX Antigenic Porphyromonas gingivalis peptides for preventing gingivitis.
 PT Claim 1; Page 563-565; 588pp; English.
 PS AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 XX Porphyromonas gingivalis (PG) polypeptide sequences given in AAX34318 to
 CC AAX34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis. (Updated on 20-MAR-2003 to correct PR field.)
 CC (Updated on 27-AUG-2003 to correct OS field.)
 XX Sequence 1232 AA;
 SQ Query Match 10.6%; Score 99; DB 2; Length 1232;
 Best Local Similarity 25.6%; Pred. No. 0.71;
 Matches 46; Conservative 23; Mismatches 47; Indels 64; Gaps 9;
 QY 16 GIDGHALADRI-----GLDEAEIATWRLSFTG-----IDDDTMAALAAEQPL---- 56
 DB 390 GLDGLASLTRLRRNQISKLEGLDRKLVRLKLDVSGNDIQSIDDIKLLAPILEQTLEKL 449
 QY 57 -----FEATADALVTDFYDHLSEYRTQDLFANSTKTVEQLKETQAEY-----LLG- 102
 DB 450 RIHDNPFVASSGILISPDYDNL-----PEIKALLEKEKEKOKTSVEYHPFCVMLLGN 503
 QY 103 --LGR-----GEYDTEYAAQARIGIKHIVLGL-----GPDVYLGAYTRYT 142
 DB 504 HSSGKTFLSQYDNTNYQ-----KNTHVLSIHRSNPNNAIFYDFGGQDYHYGIIQAFT 558
 RESULT 11
 AAY34568
 ID AAY34568 standard; protein; 1266 AA.
 XX AAY34568;
 XX 27-AUG-2003 (revised)
 DT 20-MAR-2003 (revised)
 DT 25-AUG-1999 (first entry)
 XX Porphyromonas gingivalis protein PG9.
 DE Porphyromonas gingivalis
 XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis; vaccine;
 KW antigenic.
 XX Porphyromonas gingivalis.
 OS WO9929870-A1.
 PN 17-JUN-1999.
 PD 10-DEC-1998; 98WO-AU001023.
 XX 10-DEC-1997; 97AU-00000839.
 PR 31-DEC-1997; 97AU-00001182.
 PR 30-JAN-1998; 98AU-00001546.
 PR 10-MAR-1998; 98AU-00002264.
 PR 09-APR-1998; 98AU-00002911.
 PR 23-APR-1998; 98AU-00003128.
 PR 05-MAY-1998; 98AU-00003338.
 PR 22-MAY-1998; 98AU-00003654.
 PR 29-JUL-1998; 98AU-00004917.
 PR 30-JUL-1998; 98AU-00004963.
 PR 04-AUG-1998; 98AU-00005028.
 XX


```

PD 11-SEP-1990.
XX
PF 20-JUN-1988; 88US-00209236.
XX
XX 14-FEB-1986; 86US-00829354.
PR 23-APR-1986; 86US-00854887.
PR 17-FEB-1987; 87WO-US000329.
PR 19-JUN-1987; 87US-00063959.
XX
XX (GENX ) GENEX CORP.
PA
XX Fahnstock SR;
XX
XX WPI; 1990-297491/39.
DR N-PSDB; AAQ06019.
XX
XX Recombinant Protein G variants - obt'd, using a cloned gene encoding
PT Protein G from Streptococcus sp., used for binding immunoglobulin.
XX
XX Disclosure; Fig 9; 48pp; English.
XX
XX Fragments and variants of the sequence are claimed esp. where
CC incorporated into a non-pathogenic host eg. E.coli, and expressed at high
CC levels. The variants have a higher binding efficiency and capacity for
CC immunoglobulin, and may be used for purifying, detecting and isolating
CC antibodies. (Updated on 25-MAR-2003 to correct PA field.) (Updated on 24-
CC OCT-2003 to standardise OS field)
XX
XX Sequence 593 AA;
SQ
Query Match 10.5%; Score 98; DB 2; Length 593;
Best Local Similarity 25.3%; Pred. No. 0.32;
Matches 48; Conservative 25; Mismatches 45; Indels 72; Gaps 8;
QY 10 TADVRNGIDGHALADRGIDAEIARWLSFTGIDD-----DTMAALAAE----- 53
Db 41 TPIIRNGGE---LTNLLGNSETTLARNEESATADLTAAVADTVAAAEENAGAAWEA 97
QY 54 ----QPLFEATADAL-----VTDYHLESYERTQDLPANSTKTVEQLKETQAYLLG 102
Db 98 AAAADALAKAKADALKEFNKYGVSYYKNL-----INNAKTVEGVKDLQAVV-- 145
QY 103 LGRGEVDTEYAQRARIGKHIDVLGLGPDVYLGAITYYTGLLDALADDVWADRGEEAA 162
Db 146 -----ESAKKARISEATD-----GLSDFLKSQTPA---EDTVK 175
QY 163 AVDELVARFL 172
Db 176 SIELAEAKVL 185
RESULT 14
AAR62944
ID AAR62944 standard; protein; 593 AA.
XX
XX AAR62944;
XX
XX 25-MAR-2003 (revised)
DT 10-JAN-1995 (first entry)
XX
XX Streptococcus Protein G derived from strain GX7805.
DE
XX Streptococcus Protein G; variant; IgG binding activity; immunoglobulin;
KW Lancefield Group G; bacterial Fc receptor.
XX
XX Streptococcus sp. GX7805.
OS
XX Key Location/Qualifiers
XX 106..140
FT Domain /label= A1
FT 141..178
FT Region /label= a1
FT /note= "linking region"
FT

```

```

FT Domain 179..215
FT /label= A2
FT Region 216..253
FT /label= a2
FT /note= "linking region"
FT 254..290
FT /label= A3
FT Misc-difference 269
FT /note= "corresponds to GGA codon"
FT Active-site 303..357
FT /label= B1
FT /note= "involved in IgG binding activity"
FT Region 358..372
FT /label= b
FT /note= "linking region"
FT Active-site 373..427
FT /label= B3
FT /note= "involved in IgG binding activity"
FT Misc-difference 376
FT /note= "corresponds to CCT codon"
FT Region 428..442
FT /label= b
FT /note= "linking region"
FT Active-site 443..497
FT /label= B2
FT /note= "involved in IgG binding activity"
FT Misc-difference 466
FT /note= "corresponds to ACT codon"
FT Region 531..535
FT /label= C1
FT Region 536..540
FT /label= C2
FT Region 541..545
FT /label= C3
FT Region 546..550
FT /label= C4
FT Region 551..555
FT /label= C5
FT Misc-difference 592
FT /note= "corresponds to GAA codon"
FT
FT US5312901-A.
PN
XX 17-MAY-1994.
PD
XX 21-APR-1992; 92US-00871539.
XX
XX 14-FEB-1986; 86US-00829354.
PR 23-APR-1986; 86US-00854887.
PR 19-JUN-1987; 87US-00063959.
PR 20-JUN-1988; 88US-00209236.
PR 19-JUN-1990; 90US-00540169.
XX
XX (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.
XX
XX Fahnstock SR;
XX
XX WPI; 1994-159179/19.
DR N-PSDB; AAQ75036.
XX
XX New recombinant streptococcal protein G variants - useful for antibody
PT detection and purification and for therapy.
XX
XX Example 5; Fig 9; 48pp; English.
XX
XX A 2.4kb HindIII fragment containing the entire coding sequence for
CC Protein G was isolated from Streptococcus GX7805 using the 1.9kb Protein
CC G coding sequence from Streptococcus GX7809. The Protein G has IgG-
CC binding activity which has been localised to the B repeating structure.
CC Streptococcal Protein G variants comprising the B domains are claimed.
CC (Updated on 25-MAR-2003 to correct PF field.)
XX
XX Sequence 593 AA;
SQ

```

```
Query Match      10.5%; Score 98; DB 2; Length 593;
Best Local Similarity 25.3%; Pred. No. 0.32;
Matches 48; Conservative 25; Mismatches 45; Indels 72; Gaps 8;

QY 10 TADVRNGIDGHALADRIGLDEABIAWRLSFTGIDD-----DTMAALAAE----- 53
Db 41 TPIIRNGGE---LTNLLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAAAWEA 97
QY 54 ----QPLFEATADAL-----VTDFYDHLESYERTQDLFANSTKTVBOLKETOABYLLG 102
Db 98 AAAADALAKAKADALKEFNKYGVSDYKNL-----INNAKTVEGVKDIQAQVW-- 145
QY 103 LGRGEYDTEYAAQARARIGIKIHVDVILGLGPDVILGAYTRYVTGLLDALADDVADRGEEAAA 162
Db 146 -----ESAKKARISEATD-----GLSDFLKSTPA---EDTVK 175
QY 163 AVDELVAREL 172
Db 176 SIELAEAKVL 185
```

RESULT 15
AAR10005
ID AAR10005 standard; protein; 594 AA.
XX
AC AAR10005;
XX
DT 24-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 13-MAR-1991 (first entry)
XX
DE Streptococcus GX7805 protein G.
XX Immunoglobulins; Ig.
XX
OS Streptococcus sp; GX7805.
XX
FH Key Location/Qualifiers
FT Active-site 304..358 /label= Active Site B1
FT Active-site 374..428 /label= Active Site B3
FT Active-site 444..498 /label= Active Site B2
XX
PN US4977247-A.
XX
PD 11-DEC-1990.
XX
PF 19-MAY-1989; 89US-00354264.
XX
PR 14-FEB-1986; 86US-00829354.
PR 23-APR-1986; 86US-00854887.
PR 17-FEB-1987; 87WO-US000329.
PR 19-JUN-1987; 87US-00063959.
PR 20-JUN-1988; 88US-00209236.
XX
PA (GENEX) GENEX CORP.
XX
PI Fahnestock SR, Lee T, Wroble MH;
XX
DR WPI; 1991-006758/01.
DR N-PSDB; AAR10002.
XX
PT Immobilised protein G variants - used for detection, isolation and
PT purificn. immunoglobulin(s) and immunoglobulin fragments.
XX
PS Disclosure; Fig 9; 52pp; English.
XX
CC Protein G gene product may be modified allowing the variant to be
CC immobilised and exhibit different binding profiles. The bound protein is
CC useful in purification and detection of Igs and fragments. (Updated on 25

```
CC -MAR-2003 to correct PA field.) (Updated on 25-MAR-2003 to correct PI  
CC field.) (Updated on 24-OCT-2003 to standardise OS field)  
XX  
SQ Sequence 594 AA;  
Query Match      10.5%; Score 98; DB 2; Length 594;  
Best Local Similarity 25.3%; Pred. No. 0.32;  
Matches 48; Conservative 25; Mismatches 45; Indels 72; Gaps 8;  
QY 10 TADVRNGIDGHALADRIGLDEABIAWRLSFTGIDD-----DTMAALAAE----- 53  
Db 41 TPIIRNGGE---LTNLLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAAAWEA 97  
QY 54 ----QPLFEATADAL-----VTDFYDHLESYERTQDLFANSTKTVBOLKETOABYLLG 102  
Db 98 AAAADALAKAKADALKEFNKYGVSDYKNL-----INNAKTVEGVKDIQAQVW-- 145  
QY 103 LGRGEYDTEYAAQARARIGIKIHVDVILGLGPDVILGAYTRYVTGLLDALADDVADRGEEAAA 162  
Db 146 -----ESAKKARISEATD-----GLSDFLKSTPA---EDTVK 175  
QY 163 AVDELVAREL 172  
Db 176 SIELAEAKVL 185
```

Search completed: August 10, 2004, 15:26:16
Job time : 33.8547 secs

B/ANK

Result No.	Score	Query %		Length	DB	ID	Description
		Match	%				
1	91.5	9.8		955	2	US-08-428-414A-3	Sequence 3, Appli
2	89	9.5		955	1	US-08-006-676B-1	Sequence 1, Appli
3	89	9.5		955	1	US-08-282-845-2	Sequence 2, Appli
4	89	9.5		955	5	PTC-US94-00324-1	Sequence 1, Appli
5	85.5	9.2		302	4	US-09-328-352-4846	Sequence 4846, Ap
6	83.5	8.9		319	4	US-09-489-0339A-8872	Sequence 8872, Ap
7	82	8.8		542	4	US-09-252-991A-2136	Sequence 2136, A
8	81.5	8.7		1253	4	US-09-252-991A-30019	Sequence 30019, A
9	80.5	8.6		700	4	US-09-252-991A-19384	Sequence 19384, A
10	80	8.6		553	4	US-09-252-991A-32970	Sequence 32970, A
11	80	8.6		677	4	US-09-252-991A-18102	Sequence 18102, A
12	79	8.5		438	4	US-09-540-236-2595	Sequence 2595, Ap
13	79	8.5		878	3	US-08-941-936-2	Sequence 2, Appli
14	78.5	8.4		733	4	US-09-328-352-5599	Sequence 5599, Ap
15	78.5	8.4		755	5	PTC-US93-07923-3	Sequence 3, Appli
16	78.5	8.4		759	5	PTC-US93-07923-2	Sequence 2, Appli
17	78.5	8.4		766	1	US-08-230-491A-3	Sequence 3, Appli
18	78.5	8.4		766	1	US-08-619-280A-3	Sequence 3, Appli
19	78.5	8.4		766	2	US-08-940-391-3	Sequence 3, Appli
20	78.5	8.4		766	4	US-09-794-236-1	Sequence 1, Appli
21	78.5	8.4		766	4	US-10-002-593-6	Sequence 6, Appli
22	78	8.4		443	4	US-09-328-352-6943	Sequence 6943, Ap
23	78	8.4		510	4	US-09-489-0339A-11123	Sequence 11123, A
24	77	8.3		297	4	US-09-252-991A-17482	Sequence 17482, A
25	77	8.3		320	4	US-09-252-991A-21066	Sequence 21066, A
26	76.5	8.2		579	4	US-09-252-991A-18063	Sequence 18063, A
27	76	8.1		531	4	US-08-976-063E-34	Sequence 34, Appli

OPERATING SYSTEM: Apple System 7.1
 SOFTWARE: Microsoft Word, version 5.1a
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/00324
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/006,676
 FILING DATE: 15-JAN-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Perkins, Patricia Anne
 REGISTRATION NUMBER: 34,693
 REFERENCE/DOCKET NUMBER: 5004-WO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 587-0430
 TELEFAX: (206) 233-0644
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 955 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US94-00324-1

Query Match 9.5%; Score 89; DB 5; Length 955;
 Best Local Similarity 29.3%; Pred. No. 0.3;
 Matches 49; Conservative 19; Mismatches 69; Indels 30; Gaps 7;
 QY 44 DDTMAALAEQPLFEATADALVT---DFVDHLESYERTQDLFANSTKTVQELKETOAEYL 100
 DB 546 ESTVAQLERQERREVALDQTHQKLEALESSERTA---AERDQLLQQLTELQSE-R 601
 QY 101 LGLGRGEYDTE-YAAQARIGIKHIVLGLGPDVVLGA---YTRYVTGLL----- 145
 DB 602 TQLSQVVTDRRLTRDLRLQIQEYGETELARVDVALCAQEMEARHAAVFHILQULELAT 661
 QY 146 ---DALADVADRGEAAAVDELV-----ARFLPMLKLLTFDQOI 184
 DB 662 EWEDALRERALARDEMAAELDAAASTQNARESACELTSLQQL 708

RESULT 5
 US-09-328-352-4846
 ; Sequence 4846, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 4846
 ; LENGTH: 302
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-4846

Query Match 9.2%; Score 85.5; DB 4; Length 302;
 Best Local Similarity 29.4%; Pred. No. 0.14;
 Matches 37; Conservative 23; Mismatches 41; Indels 25; Gaps 7;
 QY 25 RIGL---DEAEIATWLSFTGIDDDTMAALAEQPLFEATADALVTDVFDHLESYERTQDL 81
 DB 179 RIGLSNVDSAEILKSLP-QGPHNNITLVYL---DPPYAKGQQLYENFYNHQDHVEIMRAL 234
 QY 82 FANSTK-----TVEQLKETQAEYLLGLGRGEYDTEYAAQARIGK-----IHDVLGLG 129
 DB 235 KSSSIKWIVSYDNDVDAIRELYKDFRV-----LEVSLQTAQOKKIGEEWIFSNDRV--LI 288
 QY 130 PDVYIG 135

Db 289 PNVRLG 294

RESULT 6
 US-09-489-039A-8872
 ; Sequence 8872, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 8872
 ; LENGTH: 319
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-8872

Query Match 8.9%; Score 83.5; DB 4; Length 319;
 Best Local Similarity 25.4%; Pred. No. 0.26;
 Matches 51; Conservative 17; Mismatches 72; Indels 61; Gaps 9;
 QY 14 RRGIDGHALADRIGLDEAEIA-WRLSFTGIDDDTMAALAEQPLF-----EATADALV 65
 DB 120 RRDGVVLFGFTGIDEAMLPWR-----DTLVMARDAPGFAVCYDDEGAILTLM 171
 QY 66 TDFYDH-----LESYERTQDLFANSTKTVQELKETOAEYLLGLGRGEYDTE 111
 DB 172 QRLYDRGRHRHSFGLGVPHSDVTTGERRLAYLAFCCK-HRLTPTAALPGLGKQG-YDTV 229
 QY 112 YAAQARIGIKI---HDVLGLGPDVVLGVATRYVTGLLDAL----- 148
 DB 230 ASVLTAETSALVCATDTTALGASKYLOQOGR-----DALQASVGSVTPLMKFLHPEILT 283
 QY 149 ADDVVADRGEAAAVDELVA 169
 DB 284 VDPGVAESGRRAARQLIEQIA 304

RESULT 7
 US-09-252-991A-21396
 ; Sequence 21396, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 21396
 ; LENGTH: 542
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-21396

Query Match 8.8%; Score 82; DB 4; Length 542;
 Best Local Similarity 25.3%; Pred. No. 0.84;
 Matches 49; Conservative 19; Mismatches 66; Indels 60; Gaps 8;
 QY 3 NDNDTLVTADVRNGIDGHADRIGLDEAEIATWLSFTGIDDDTMAALAEQPLFEATAD 62
 DB 149 NENDTVVTDEIRFG-----DNDTAAALVAN--LVEADLL 180

```
QY 63 ALVTD---FYDHLSEYRTQDLFANSTKTVQKETAQYLLGLGRGEYDTEYAAQR--A 117
Db 181 VILTRDGMFDAPRNNPDQALIYERADDPQLDVAAGSAGALGCGMOTKLRAARLAA 240
QY 116 RIGKIHDLVGLGPDVYLGAYTRYTGLLDALADDVVA-----DRGEEAAAADDEL 167
Db 241 RSG-----GHTVIVG-----GRIEVLDRAGERLGLTLPDRSRKAAR--KQW 283
QY 168 VARFLPMLKLLTFD 181
Db 284 LAGHLQMRGTLVLD 297

RESULT 8
US-09-252-991A-30019
; Sequence 30019, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30019
; LENGTH: 1253
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30019

Query Match 8.7%; Score 81.5; DB 4; Length 1253;
Best Local Similarity 22.5%; Pred. No. 3.4;
Matches 48; Conservative 31; Mismatches 75; Indels 59; Gaps 11;

QY 2 SMDNDTLVTAD-VRNGIDGHALADRIGLDEAEIARLSTFTGIDDDTMAALAAEQPLFEAT 60
Db 1024 SGDHDLQHAFLAEHALDGE--DRRLGVGEV-----DGLDQDVGA-----AFDQA 1068
QY 61 ADALVTDFDYDHLSEYRTQDLF-----ANSTKTVQKETAQYLLGLG 104
Db 1069 AGRLDVVLHQFVGEVDVAVAGVHVHVRGNRAGAAAGRAEHAGDEARLVRLVGLGLRVRHLAQQA 1128
QY 105 RGSYDTEYAAQRAIRIGKIHDLVGLG-----PDVYLGAYTRYTGLLDALADDVVA 154
Db 1129 R-PFEVEFVQQR-----LHNVGLGHLGGVGVGLGVLEVDVAGV-----EVGLLDGL-DHVRA 1177
QY 155 DRGEEAAA-----VDELVARFLPMLKLLTFD 181
Db 1178 AQOQEVVAFVHVARPVGEALAAVVLGFLQVLD 1210

RESULT 9
US-09-252-991A-19384
; Sequence 19384, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
```

```
; SEQ ID NO 19384
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19384

Query Match 8.6%; Score 80.5; DB 4; Length 700;
Best Local Similarity 24.6%; Pred. No. 1.8;
Matches 49; Conservative 21; Mismatches 82; Indels 47; Gaps 8;

QY 12 DVNRIGIDGHALADRIGLDEAEIARLSTFTGIDDDTMAALAAEQPLFEATADALV---TDF 68
Db 129 DARRPADAHVPVVGQRAF-QATLAQR-----QVTVATAAGQRLAALAAAGDVQAQHTQY 180
QY 69 YDHLSEYRTQDLFANSTKTVQKETAQYLLGLGRGEYDTEYAAQRAIRIGKI-HDVLG 127
Db 181 AEHQDQDEQHAHLLDDVPEPLRVERNAVDFLGLGRGERRVDHADAVAAAGVVAHRLV- 239
QY 128 LGPDVYLGAYTRYTGLLDALAD-----DV-----VAERGE- 159
Db 240 -----YQAGEARQFVGAPGLVADLAGLRVDLHLVVQHHRRHRSQAQASILLLLGVADRAVQF 294
QY 160 -AAAADVELVARFLPMLKL 177
Db 295 VVAGGLVALVGRVRRQLRL 313

RESULT 10
US-09-252-991A-32970
; Sequence 32970, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32970
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32970

Query Match 8.6%; Score 80; DB 4; Length 553;
Best Local Similarity 21.1%; Pred. No. 1.5;
Matches 41; Conservative 28; Mismatches 79; Indels 46; Gaps 5;

QY 20 HALADRI-----GLDEAEIARLSTFTGIDDDTMAALAAEQPLFEATADALVTDFYDHL 74
Db 137 HAPAPRLLSAFPLDAPAVGLGPAVEGLAVDEYVAQAQGRPLRSFFFAESAEQKVIHRAA 196
QY 75 YERTQDLFANSTKTVQKETAQY-----YLLGLGRGEYDTEYAAQ-----RARIGK-- 121
Db 197 LHALQQLDGLVLRGDAQAPAHQVEVVVGLVEGEQVEAGAQFVHFHQRQLCGGG 256
QY 122 -----IHDVL-----GLGPDVYLGAYTRYTGLLDALADDV 153
Db 257 QVPMGDRYLVAEGVAAAGVGVVADVIGVEIVEBGVGVVVEGNAEDHRVVGVHHHPVAEIG 316
QY 154 ADGEEAAAADDEL 167
Db 317 LPAGDEFGVALDDL 330

RESULT 11
US-09-252-991A-18102
; Sequence 18102, Application US/09252991A
```

```

: Patent No. 6551795
:
: GENERAL INFORMATION
:
: APPLICANT: Marc J. Rubenfield et al.
:
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
:
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
:
: FILE REFERENCE: 107196.136
:
: CURRENT APPLICATION NUMBER: US/09/252,991A
:
: CURRENT FILING DATE: 1999-02-18
:
: PRIOR APPLICATION NUMBER: US 60/074,788
:
: PRIOR FILING DATE: 1998-02-18
:
: PRIOR APPLICATION NUMBER: US 60/094,190
:
: PRIOR FILING DATE: 1998-07-27
:
: NUMBER OF SEQ ID NOS: 33142
:
: SEQ ID NO 18102
:
: LENGTH: 677
:
: TYPE: PRT
:
: ORGANISM: Pseudomonas aeruginosa
:
: US-09-252-991A-18102

```

Query Match	8.6%;	Score 80;	DB 4;	Length 677;	
Best Local Similarity	26.6%;	Pred. No. 2;			
Matches	45;	Conservative 18;	Mismatches 64;	Indels 42;	Gaps 10;

QY	6	DTIATVDVRNGIDGHALADRGIDLEARIWLSFTGIDDDTMAALAAEQPLFEATADALV	65
		: : :	
Db	245	DKLUSGDTTPMADEVRDELNAR-----SWR-----AFIGDFFPLSRREDREALI	288
		: : :	
QY	66	TDFYDHLESYERTQDLFANSTKTKVEQ-----LKETOAEYLL-GLGRGEYDTEVAQAQRARI	119
		: : :	
Db	289	A-----LYESPRDYLAG--KSVEEKETVLAKTSTRDYLLKNVLGSETSVKVPQ-----	334
		: : :	
QY	120	GKIHDLVLGPDVYVLGAYTRYTGTGL--LDALADDVADRGEAAAAAYDE	166
		: : :	
Db	335	GRSNDFSALGADA-LPAADAYAGFPFGFDALG---LPQPSGEACAEWDE	379
		: : :	

```

RESULT 12
US-09-540-236-2595
; Sequence 2595, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2595
; LENGTH: 438
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-2595

```

[illegible]

```

RESULT 13
US-08-941-936-2
; Sequence 2, Application US/08941936
; Patent No. 6054305
; GENERAL INFORMATION:
; APPLICANT: Tatsumi, Hiroki
; APPLICANT: Etsaki, Maoki
; APPLICANT: Horiuchi, Tatsuo
; APPLICANT: Nagahara, Ayumu
; TITLE OF INVENTION: Pyruvate Orthophosphate Dikinase Gene,
; TITLE OF INVENTION: Recombinant DNA, and Process For Producing Pyruvate
; TITLE OF INVENTION: Orthophosphate Dikinase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/941.936
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: HIRAKI-03009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-705-8410
; TELEFAX: 415-397-8338
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 878 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-941-936-2

```

Query Match	8.5%;	Score 79;	DB 3;	Length 878;
Best Local Similarity	25.3%;	Pred. No. 3..9;		
Matches	49;	Conservative 26;	Mismatches 67;	Indels 52; Gaps 10
QY	22	LADRLGLDEAEIAWLSFTGID-----DDTMAALAAEQ-----	PLPFAATADA--	LVTD 67
Ddb	329	LQTVGKRTAEAPRIATQVLDGLTMDAEAVARVTGDLQALMPFPAATAADARLLTTG	388	
QY	68	FY-----DHLESYERTODLFANSTKTEQKQTAQAEYLLGL--	GRGEY-----	DTEY 112
Ddb	389	MNASPGAAVGKAVFSSERAVELAGQGEAVILVRREINPDDLAGMAAARGVLTSGGKTS	448	
QY	113	AAQPARIGKIHVLGPGVYLCA-----YTRYT--GLLDALADDVVADRGEEAAA	163	
Ddb	449	AAVVAR-----GMKTCVCGAELEVPDHPARRFTAFGGIVVNGEVEISIDGSTGAVY	500	
QY	164	VDEL-----VARF 171		
Ddb	501	LGEVPTTASPVARY 514		

RESULT 14
US-09-328-352-5599
; Sequence 5599, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.

```
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5599
; LENGTH: 733
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5599

Query Match      8.4%; Score 78.5; DB 4; Length 733;
Best Local Similarity 21.6%; Pred. No. 3.4; 74; Indels 83; Gaps 10;
Matches 52; Conservative 32; Mismatches 74; Indels 83; Gaps 10;

QY 9 VTADVNRGIDGHALADR-----IGLDEAEIAWRISFTGIDDD-----TMAALA 51
Db 223 IFAVDNITSSNSVAERGKLTGILG-----LNYQGYDKHEHITQVLNAILVTYGAQN 273
QY 52 AEQPLFEAT-----ADALVTDIFYDHLESYERTQDLPANSTKTYE-----90
Db 274 VERRAESAQTLKFLDEQLPDLKKQLDDAEROFNKFROQYNTVDVTKESELYLTQSITILE 333
QY 91 ----OLKETQAEVLLGLGRGEYDTEYAAOR-----ARIGKIHVDVLGLGPDV---Y 133
Db 334 TKKAELEQQAEM-----AAKYTAEPHFWREINGQITAINKQIGELNSTLKQLPDVQRY 388
QY 134 LGAY-----TRYVTGLLDALADDVVADRGE-----EAAAADVELVARFLPMIKLIT 179
Db 389 LQLYREVEVKTQLYTALLNSYQQLRIAKAGELGNVRIVDTAVEPVEPIKPKKLQVLISL 448
QY 180 F 180
Db 449 F 449
```

```
RESULT 15
PCT-US93-07923-3
; Sequence 3, Application PC/TUS9307923
; GENERAL INFORMATION:
; APPLICANT: Morimoto, Chikao
; APPLICANT: Schlossman, Stuart F.
; APPLICANT: Tanaka, Toshiaki
; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: Wordperfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07923
; FILING DATE: 19930819
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/934,162
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: 07/832,211
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/055002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
```

```
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 755
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
PCT-US93-07923-3

Query Match      8.4%; Score 78.5; DB 5; Length 755;
Best Local Similarity 25.8%; Pred. No. 3.5;
Matches 47; Conservative 25; Mismatches 73; Indels 37; Gaps 8;

QY 2 SNDNDTLVTADVR-NGIDG---HALADRIGL---DEAEIAWRISFTGIDDD-----45
Db 558 STENIIIVASFDRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQFSKMGFVDNKRIAIWG 617
QY 46 -----TMAALAAEQPLFE---ATA-----DALVTDFYDHLESYERTQDLPANSTK 87
Db 618 WSYGGVVTSMVLGSGGVFKCGIAVAPVSRWEYDVSVYTERYMGLPTPEDNLDHYRNSTV 677
QY 88 TVEQLKETQAEVLLGLGRGEYDTEYAAQPARIGKIHVDVLGLGPDVYVLGAYTRYTGLLDA 147
Db 678 MSRAENFKQVEYLLIHGTAD-DNVHFQSSAQISKA--LVDVGVDFOAMWYTDDEHGIASS 734
QY 148 LA 149
Db 735 TA 736
```

Search completed: August 10, 2004, 15:30:20
Job time : 9.35593 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:28:46 ; Search time 26.7312 Seconds
(without alignments)
2159.184 Million cell updates/sec

Title: US-09-455-978B-77

Perfect score: 933

Sequence: 1 MSNDNDTLVTADVRNGIDGH.....DELVARFLPMLKLLTFDQOI 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	99	10.6	448	12	US-10-460-524-5
2	94.5	10.1	883	15	Sequence 5, Appli
3	94	10.1	218	14	Sequence 18563, A
4	89.5	9.6	400	14	Sequence 9281, Ap
5	88	9.4	505	12	Sequence 13480, A
6	87	9.3	496	12	Sequence 66025, A
7	86	9.2	362	16	Sequence 6364, A
8	85	9.1	2703	12	Sequence 137315, A
9	83.5	8.9	258	14	Sequence 66108, A
10	83.5	8.9	315	12	Sequence 11335, A
11	83.5	8.9	367	12	Sequence 59961, A
12	83.5	8.9	368	12	Sequence 72593, A
13	83.5	8.9	1254	12	Sequence 48921, A
14	83	8.9	384	12	Sequence 53778, A
15	82.5	8.8	403	12	Sequence 48636, A
					Sequence 61978, A

16	82.5	8.8	695	12	US-10-282-122A-69832	Sequence 69832, A
17	82.5	8.8	765	15	US-10-369-493-13529	Sequence 13529, A
18	82.5	8.8	831	9	US-09-738-626-5468	Sequence 5468, Ap
19	82.5	8.8	928	12	US-10-282-122A-61748	Sequence 61748, A
20	82.5	8.8	939	14	US-10-156-761-12605	Sequence 12605, A
21	82	8.8	318	12	US-10-425-114-63899	Sequence 63899, A
22	82	8.8	329	12	US-10-289-456-16	Sequence 16, Appl
23	82	8.8	329	12	US-10-622-064-28	Sequence 16, Appl
24	82	8.8	329	14	US-10-243-739-16	Sequence 28, Appl
25	82	8.8	329	14	US-10-244-065-16	Sequence 16, Appl
26	82	8.8	329	14	US-10-289-454-16	Sequence 16, Appl
27	82	8.8	329	15	US-10-346-190-16	Sequence 16, Appl
28	82	8.8	329	15	US-10-465-811-7	Sequence 7, Appl
29	82	8.8	329	16	US-10-622-124-10	Sequence 10, Appl
30	82	8.8	329	16	US-10-622-087-10	Sequence 10, Appl
31	82	8.8	330	14	US-10-050-902-254	Sequence 254, App
32	82	8.8	330	14	US-10-050-898-254	Sequence 254, App
33	82	8.8	901	15	US-10-369-493-18812	Sequence 18812, A
34	82	8.8	1333	12	US-10-282-122A-55546	Sequence 55546, A
35	82	8.8	2799	12	US-10-128-714-8240	Sequence 8240, A
36	81	8.7	1178	14	US-10-128-714-8240	Sequence 8240, A
37	80.5	8.6	546	15	US-10-369-493-8893	Sequence 8893, Ap
38	80.5	8.6	1173	15	US-10-369-493-11290	Sequence 11290, A
39	80.5	8.6	2132	12	US-10-424-599-161362	Sequence 161362, A
40	80.5	8.6	2151	12	US-10-424-599-161366	Sequence 161366, A
41	80.5	8.6	2221	12	US-10-424-599-161366	Sequence 161366, A
42	80	8.6	811	16	US-10-437-963-171424	Sequence 171424, A
43	80	8.6	954	12	US-10-282-122A-69774	Sequence 69774, A
44	79.5	8.5	429	12	US-10-282-122A-62133	Sequence 62133, A
45	79.5	8.5	1015	16	US-10-437-963-184684	Sequence 184684, A

ALIGNMENTS

RESULT 1

US-10-460-524-5
; Sequence 5, Application US/10460524
; Publication No. US20040029781A1
; GENERAL INFORMATION:
; APPLICANT: Hernan, Ronald A
; APPLICANT: Mehig, Richard J
; APPLICANT: Brockie, Ian
; APPLICANT: Jenkins, Elizabeth
; TITLE OF INVENTION: Affinity Peptides and Method for Purification of Recombinant Proteins
; FILE REFERENCE: SGM 7047.1
; CURRENT APPLICATION NUMBER: US/10/460.524
; CURRENT FILING DATE: 2003-06-12
; PRIOR APPLICATION NUMBER: US 60/388,059
; PRIOR FILING DATE: 2002-06-12
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Streptococcus
US-10-460-524-5

Query Match 10.6%; Score 99; DB 12; Length 448;
Best Local Similarity 25.3%; Pred No. 0.087; Indels 72; Gaps 8;
Matches 48; Conservative 25; Mismatches 45;
Qy 10 TADVRNGIDGHADRLGRIGLDEAEIAWRLSPTGIDD-----DTMAALAAE----- 53
Db 41 TPIIRNGE---LTNLLGNSETTLALRNEESATADLTAAAVADTVAAAAENAGAAWEA 97
Qy 54 -----QPIFETADAL-----VTFDYHLESYERTQD.FANSTKTVQKETAQAEYLIG 102
Db 98 AAAADALAKAKADALKEFNKYGVSDYKXNL-----INNAKTVEGKIDYQAQVW-- 145
Qy 103 LGRGEVDTEVAQAQRARIGKIHVDVLGIPDVLGAYTRYTGLLDALADVDVADRGEEAAA 162
Db 146 -----ESAKKARISATD-----GLSDFLKSTQTPA---EDTVK 175

QY	Db	163	AVDELVARFL 172	10.1%; Score 94; DB 14; Length 218;	Best Local Similarity 28.7%; Pred. No. 0.11; Matches 47; Conservative 22; Mismatches 73; Indels 22; Gaps 8;
QY	Db	176	STELAEAKVL 185	10.1%; Score 94; DB 14; Length 218;	Best Local Similarity 28.7%; Pred. No. 0.11; Matches 47; Conservative 22; Mismatches 73; Indels 22; Gaps 8;
<p>RESULT 2</p> <p>US-10-369-493-18563</p> <p>; Sequence 18563, Application US/10369493</p> <p>; Publication No. US20030233675A1</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: Cao, Yongwei</p> <p>; APPLICANT: Hinkle, Gregory J.</p> <p>; APPLICANT: Slater, Steven C.</p> <p>; APPLICANT: Goldman, Barry S.</p> <p>; APPLICANT: Chen, Xianfeng</p> <p>; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF</p> <p>; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES</p> <p>; FILE REFERENCE: 38-10(52052)B</p> <p>; CURRENT APPLICATION NUMBER: US/10/369,493</p> <p>; CURRENT FILING DATE: 2003-02-28</p> <p>; PRIOR APPLICATION NUMBER: US 60/360,039</p> <p>; PRIOR FILING DATE: 2002-02-21</p> <p>; NUMBER OF SEQ ID NOS: 47374</p> <p>; SEQ ID NO 18563</p> <p>; LENGTH: 883</p> <p>; TYPE: PRT</p> <p>; ORGANISM: Halobacterium sp. NRC-1</p> <p>US-10-369-493-18563</p>					
QY	Db	12	DVRNGIDGH--ALADRIGLDEAEI--AWRLSF--TGIDDDTMAALAAEQPLFEATADALVTDFYDHLESVER 77	10.1%; Score 94.5; DB 15; Length 883;	Best Local Similarity 28.3%; Pred. No. 0.67; Mismatches 23; Indels 43; Gaps 10;
QY	Db	180	DVKSNNVEGQLDRDLAQIA--DKEAADPHDLASHNTALAEVTADEHFEAREQARQTRD 237	10.1%; Score 94.5; DB 15; Length 883;	Best Local Similarity 28.3%; Pred. No. 0.67; Mismatches 23; Indels 43; Gaps 10;
QY	Db	63	ALVTDFYDHLESVERTQDLFANSTKTVEQLKETOAYLILGLGRGEVDTEYAAQRA-----TAD 62	10.1%; Score 94.5; DB 15; Length 883;	Best Local Similarity 28.3%; Pred. No. 0.67; Mismatches 23; Indels 43; Gaps 10;
QY	Db	238	----DAADVLYEESRTALADVEETIADVREAAE-----AERERTLADRVSDHRE 286	10.1%; Score 94.5; DB 15; Length 883;	Best Local Similarity 28.3%; Pred. No. 0.67; Mismatches 23; Indels 43; Gaps 10;
QY	Db	118	RIGKIHVV-----LGL-GPDVVLGAYTRYTGLLDALAD--DVVADRGEAAAAVDEL 167	10.1%; Score 94.5; DB 15; Length 883;	Best Local Similarity 28.3%; Pred. No. 0.67; Mismatches 23; Indels 43; Gaps 10;
QY	Db	287	RASLDDEAALAAADLGLDDPDAEDASAER-----DAVADQEEAAVREVPVAVSEL 340	10.1%; Score 94.5; DB 15; Length 883;	Best Local Similarity 28.3%; Pred. No. 0.67; Mismatches 23; Indels 43; Gaps 10;
<p>RESULT 3</p> <p>US-10-156-761-9281</p> <p>; Sequence 9281, Application US/10156761</p> <p>; Publication No. US20030119018A1</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: OMURA, SATOSHI</p> <p>; APPLICANT: IKEDA, HARUO</p> <p>; APPLICANT: ISHIKAWA, JUN</p> <p>; APPLICANT: HORIKAWA, HIROSHI</p> <p>; APPLICANT: SHIBA, TADAYOSHI</p> <p>; APPLICANT: SAKAKI, YOSHIYUKI</p> <p>; APPLICANT: HATTORI, MASAHIRA</p> <p>; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES</p> <p>; FILE REFERENCE: 249-262</p> <p>; CURRENT APPLICATION NUMBER: US/10/156,761</p> <p>; CURRENT FILING DATE: 2002-05-29</p> <p>; PRIOR APPLICATION NUMBER: JP 2001-204089</p> <p>; PRIOR FILING DATE: 2001-05-30</p> <p>; PRIOR APPLICATION NUMBER: JP 2001-272697</p> <p>; PRIOR FILING DATE: 2001-08-02</p> <p>; NUMBER OF SEQ ID NOS: 15109</p> <p>; SEQ ID NO 9281</p> <p>; LENGTH: 218</p> <p>; TYPE: PRT</p> <p>; ORGANISM: Streptomyces avermitilis</p> <p>US-10-156-761-9281</p>					
QY	Db	42	IDDDTMAALAAEQPLFEATADALVTDFYDHLESY--ERTQDLFANSTKTVEQLKETOAY 99	10.1%; Score 94.5; DB 15; Length 883;	Best Local Similarity 28.3%; Pred. No. 0.78; Mismatches 16; Indels 17; Gaps 5;
QY	Db	2	LSEQSAATVRATLPAVGAAVGEITARFYDLRFAARPELLRDLFNRG-----NOAAGTQROA 57	10.1%; Score 94.5; DB 15; Length 883;	Best Local Similarity 28.3%; Pred. No. 0.78; Mismatches 16; Indels 17; Gaps 5;
QY	Db	100	LLG--LGRGEVDTEYAAQRA-----RIGKIHVILGLGPDVVLGAYTRYTGLLDALADDV 152	10.1%; Score 94.5; DB 15; Length 883;	Best Local Similarity 28.3%; Pred. No. 0.78; Mismatches 16; Indels 17; Gaps 5;
QY	Db	58	LASIAAPATYLVVHPDPDPDAMDRIAHKHSLSGIAPGQYAVVHHLFAAIAEVLGDV 117	10.1%; Score 94.5; DB 15; Length 883;	Best Local Similarity 28.3%; Pred. No. 0.78; Mismatches 16; Indels 17; Gaps 5;
QY	Db	153	VADRGEAAAAVDEL 167	10.1%; Score 94.5; DB 15; Length 883;	Best Local Similarity 28.3%; Pred. No. 0.78; Mismatches 16; Indels 17; Gaps 5;
QY	Db	118	T-----PEVAAWDEV 128	10.1%; Score 94.5; DB 15; Length 883;	Best Local Similarity 28.3%; Pred. No. 0.78; Mismatches 16; Indels 17; Gaps 5;
<p>RESULT 5</p> <p>US-10-282-122A-66025</p> <p>; Sequence 66025, Application US/10282122A</p> <p>; Publication No. US20040029129A1</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: Wang, Lianguo</p> <p>; APPLICANT: Zamudio, Carlos</p> <p>; APPLICANT: Malone, Cheryl</p> <p>; APPLICANT: Haselbeck, Robert</p> <p>; APPLICANT: Ohlsen, Kari</p> <p>; APPLICANT: Zyskind, Judith</p> <p>US-10-282-122A-66025</p>					

```

; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66025
; LENGTH: 505
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-10-282-122A-66025

Query Match
Best Local Similarity 9.4%; Score 88; DB 12; Length 505;
Matches 46; Conservative 24; Mismatches 69; Indels 40; Gaps 7;

QY 6 DTLVTADVNRGIDGHADRLGDEAEIAWRLSP---TGIDDDTWAALAAEQPLF----- 57
Db 325 DQALAIAGRGQNVRLASDLTG-----WQINIMTSAEDERNAEDAARLRFMDHLN 377
QY 58 --EATADALVTDFDHLR--SVERTQDLFANSTKTVEQLKETOAEYLLGLGRGEYDTBYA 113
Db 378 VDEETADVIVQEGFATLEEVAVPAAELLA-----IEGFDEEIVDMLRNRARDAILTWAI 432
QY 114 AORARIGKIHDLV---GLGPDVYLGAITYYTGILLDALADDVVDRCGEAAAVDELV 168
Db 433 AAEEKLGEVSDMRNLEGIDAD-----MLRSLAEGITTRDDLAELAVDELI 479

RESULT 6
US-10-282-122A-65364
; Sequence 65364, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

```

```

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 65364
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-282-122A-65364

Query Match
Best Local Similarity 25.7%; Score 87; DB 12; Length 496;
Matches 46; Conservative 24; Mismatches 69; Indels 40; Gaps 7;

QY 6 DTLVTADVNRGIDGHADRLGDEAEIAWRLSP---TGIDDDTWAALAAEQPLF----- 57
Db 316 DRLALAIAGRGQNVRLASDLTG-----WQINIMTSAEDERNAEDAARLRFMDHLN 368
QY 58 --EATADALVTDFDHLR--SVERTQDLFANSTKTVEQLKETOAEYLLGLGRGEYDTBYA 113
Db 369 VDEETADVIVQEGFATLEEVAVPAAELLA-----IEGFDEEIVDMLRNRARDAILTWAI 423
QY 114 AORARIGKIHDLV---GLGPDVYLGAITYYTGILLDALADDVVDRCGEAAAVDELV 168
Db 424 AAEEKLGEVSDMRNLEGVDADMLL-----SLAEGITTRDDLAELAVDELI 470

RESULT 7
US-10-437-963-137315
; Sequence 137315, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 137315
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Oryza sativa

```

```
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_3880C.1.pcp
US-10-437-963-137315

Query Match          9.2%; Score 86; DB 16; Length 362;
Best Local Similarity 26.4%; Pred. No. 1.6;
Matches 51; Conservative 20; Mismatches 78; Indels 44; Gaps 9;

QY 8 LVTADVRNGIDGHADRIGLDEABIAWR-----LSFT--GIDDDTMAALAAPOQLFEATA 61
DB 11A:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 61
DB 26 ILADESTGTIGKRLAS-IGVENVEENRRALRELLFTAPGALDCLSGVILFEETLYQSTR 84
QY 62 DALVTDFVDHLESY-----ERTQDLFANSTYTVQLKETQAEYLLGLGRGEYDTF 112
DB 85 DG--TPFVDVLAAGVLAGIKVDKGTVELAGTDRFTTQGH-----GLGRCRRY 133
QY 113 AACRARIKIHVDVLGLG-----PDVILGAYTRY-----YTGLLDALADDVVAD-- 155
DB 134 YAAGARFAKRAVLISGRASSRPSQLAVDANNAQGLARVAIIQCENGLVPIVEPEILVDGE 193
QY 156 RGEERAAAADVELV 168
DB 194 HGIEACAETVTRV 206

RESULT 8
US-10-282-122A-66108
; Sequence 66108, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66108
; LENGTH: 2703
; TYPE: PRT
; ORGANISM: Neisseria meningitidis

US-10-282-122A-66108
Query Match          8.9%; Score 83.5; DB 14; Length 258;
Best Local Similarity 24.1%; Pred. No. 1.9;
Matches 49; Conservative 21; Mismatches 78; Indels 55; Gaps 9;

QY 8 LVTADVRNGIDGHADRIGLDEABIAWRLSFTGIDDDTMAALAAPOQLFEATADALVTD 67
DB 33 VVADVLDLDD--QGEALAKEIGARYVHLD-----VGREDDWQAAV-----TVAKD 74
QY 68 FVDHLESY-----RTQDLFANSTYTVQLKE--TQAEYLLGLGRGEYDTEYAAQPARIG 120
DB 75 AYGHIDGLVNNAGILRFNDLVGTPLAABFQOIVQVNVGVFLGIKTVAPIE--AAGGTIV 133
QY 121 KIHVDVLGLGPDVYVIGAYT-----RYTGLLD--ALADDVVA 154
DB 134 NTASYAGLTGMAYVGYATTKAIVGLTRVALELAACKIRVNACPGDAITAMSPSQL 193
QY 155 DRG---EAAAAVDELVARFLPM 174
DB 194 DPGADPEETARALSELVGRVLP 216

RESULT 10
US-10-282-122A-59961
```



```

; Sequence 59961, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59961
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-59961

Query Match
Best Local Similarity 8.9%; Score 83.5; DB 12; Length 315;
Matches 51; Conservative 17; Mismatches 72; Indels 61; Gaps 9;

QY 14 RRGIDGHADRIGLDEAEIA-WPLSFTGIDDDTMAALAAEQPLF-----ETADALV 65
Db 116 RRNDGVVLFGTGDIDAMLAAPWR-----DTLVLMARDAPGFASVCYDDEGAITLM 167
QY 66 TDFYDH-----LESYRTQDLFANSTKTVQELKETQAEYLLGLGR-----R 116
Db 168 QRLYDRGRHHSIFLGVPHSDVTTGERRHLYAFCKK-HRLTPTAALPGLGMKQG-YDVT 225
QY 112 YAAQARIGKI---HVLGIGPVDYVLGAYTRYTYTGLLDAL----- 148
Db 226 ASVLTAEFTSALVCATDTLTALGASKYLQOQGR-----DALQLASVGTSLPKFLHPIIT 279
QY 149 ADDVADRGEAAAVDELVA 169
Db 280 VDPGYAESGRRAKQLIEQIA 300

RESULT 11
US-10-425-114-72593
; Sequence 72593, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48921
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: CC-ZMP0148032B07_FLI.pep
US-10-425-114-48921

Query Match
Best Local Similarity 8.9%; Score 83.5; DB 12; Length 368;
Matches 47; Conservative 24; Mismatches 68; Indels 47; Gaps 10;

QY 16 GIDGHADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFETADALVTDYFDHLESY 75
Db 152 GVDGR---RVGLDEEAPA-----DADAVAAQHRDHLDD--VEALLVELVDLGH- 196
QY 76 ERTQDLFANSTKTV-----EQKETQAEYLLGLGR-----GEYDTEVAAG-----R 116
Db 197 ---RDVRAQVGVVLAVGHERVAPPQGHVILGPARQLHRVAGAHRDDVRAHHVRALLIQ 253
QY 117 ARIGKIHVDVLGIPDV---YLGAYTRY---VTGLDLDADDVVDARGEEAAAVDELVA 169
Db 254 LRLDLDLHVHVLGVDVGAGLLGLVRRVVDQDGRVAICAVVEDEAERAGGAG---V 310
QY 170 RFLPML 175
Db 311 SFEPLL 316

RESULT 12
US-10-425-114-48921
; Sequence 48921, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48921
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: CC-ZMP0148032B07_FLI.pep
US-10-425-114-48921

Query Match
Best Local Similarity 8.9%; Score 83.5; DB 12; Length 368;
Matches 47; Conservative 24; Mismatches 68; Indels 47; Gaps 10;

QY 16 GIDGHADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFETADALVTDYFDHLESY 75

```

```

; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 72593
; LENGTH: 367
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: CC-ZMP014820017B02_FLI.pep
US-10-425-114-72593

Query Match
Best Local Similarity 8.9%; Score 83.5; DB 12; Length 367;
Matches 47; Conservative 24; Mismatches 68; Indels 47; Gaps 10;

QY 16 GIDGHADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFETADALVTDYFDHLESY 75
Db 152 GVDGR---RVGLDEEAPA-----DADAVAAQHRDHLDD--VEALLVELVDLGH- 196
QY 76 ERTQDLFANSTKTV-----EQKETQAEYLLGLGR-----GEYDTEVAAG-----R 116
Db 197 ---RDVRAQVGVVLAVGHERVAPPQGHVILGPARQLHRVAGAHRDDVRAHHVRALLIQ 253
QY 117 ARIGKIHVDVLGIPDV---YLGAYTRY---VTGLDLDADDVVDARGEEAAAVDELVA 169
Db 254 LRLDLDLHVHVLGVDVGAGLLGLVRRVVDQDGRVAICAVVEDEAERAGGAG---V 310
QY 170 RFLPML 175
Db 311 SFEPLL 316

```

Db 152 GVDR-----RVGLDEEAPA-----DADAVAAQHRDHLDD--VEALLVELVDELGH- 196
QY 76 ERTQDLFANSTKTV-----EQKETOAEYLLGLGR-----GEYDTEVAAQ-----R 116
Db 197 ---RDVRAQVGVGLAVGHVAPPPQGHVGLGPQRLHRVAGAHRDDVVRARHVRALLLQ 253
QY 117 ARIGKIHVDLGLGPDV-----VLGYATRY---YTGLLDALDDVVDVDRGEEAAAVDELVA 169
Db 254 LRLDLHDVHALHGVAGLLGLGLVHRVHRVQDGRVAAICBAVVEDEAEAGGGAG---V 310
QY 170 RFLPML 175
Db 311 SFEPLL 316

RESULT 13

US-10-282-122A-53778

; Sequence 53778, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 53778

; LENGTH: 1254

; TYPE: PRT

; ORGANISM: *Corynebacterium diptheriae*

US-10-282-122A-53778

Query Match 8.9%; Score 83.5; DB 12; Length 1254;
Best Local Similarity 25.08; Pred. No. 16;
Matches 52; Conservative 23; Mismatches 82; Indels 51; Gaps 10;

QY 4 DNDTLV-----TAVRNSIDGHALADIGLDE-----REIAWRLSFTG 41
Db 658 DNLNLVNHVALFDPNEDLRNAADGAGYEDDITLDEFVNTAPQMRFLGVKKAWELR-DG 716

QY 42 IDDDTMAALAAEQFLFATADALVTDYDHLSEYERTQDLFANSTKTVQKETAQAEYLL 101
Db 717 AGPQVQA--AADSALKGTGTPDAL--DEFVNGDGYEKAR--YLDQVQQAAYELTDT----- 764
QY 102 GLGRGEYDTEYAAQARARIGKIHVDLGLGPDVVLGYATRYTGLLDALAD-----DVVAD 155
Db 765 ---GGPEVQTAEEAAVTG---DRQQLDEFVSIQGYRR---AILDORSORDAHNAEINALLS 814
QY 156 RGEAAAAVDELVARFLPMLKLLTFFDQ 183
Db 815 AQONAAELASQEAANAQEAATRAATGDAQ 842

RESULT 14

US-10-425-114-48636

; Sequence 48636, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E.

; APPLICANT: Tabaska, Jack E.

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 48636

; LENGTH: 384

; TYPE: PRT

; ORGANISM: *Zea mays*

; FEATURE:

; OTHER INFORMATION: Clone ID: LIB31116-002-B8_FLI.pep

US-10-425-114-48636

Query Match 8.9%; Score 83; DB 12; Length 384;
Best Local Similarity 25.6%; Pred. No. 3.6;
Matches 31; Conservative 25; Mismatches 43; Indels 22; Gaps 5;

QY 58 EATADALVTDF-YDHLSEYERTQDLFANSTKTVQKLE-----TQAEYLLGLGRGEYD 109
Db 267 KANSAPITAYRYHSLCSYMGDDMF-SSDLSQDLRQLGHMSITQCQVIFSMG-DEYV 324
QY 110 TEYAAQARARIGKIHVDLGLGPDVVLGYATRYTGLLDALADDVADRGEEAAAVDELVA 169
Db 325 PEYVDKRALDLRLCALGGAKEVEI-----EWGNHLSNVEQEAIVVDFVK 372
QY 170 R 170
Db 373 R 373

RESULT 15

US-10-282-122A-61978

; Sequence 61978, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

```
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61978
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Mycobacterium avium
US-10-282-122A-61978

Query Match      8.8%; Score 82.5; DB 12; Length 403;
Best Local Similarity 25.4%; Pred. No. 4.4;
Matches 36; Conservative 20; Mismatches 65; Indels 21; Gaps 6;

QY 42 IDDDTMAALAAEQPLEFATADALVTDYDHL--ESYERTQDLFANSTKTVEQLKETQAEY 99
Db 25 ISDSVLDALLAQDPRSRVAVETLVTGQVHVGEVTTTAKAFADITNTVR-----ER 77

QY 100 LLGLGRGEVDTETAYAAQRARIGKIHVDVLGLG---PDVYLGAHYTRYT---GLLDALADDVV 153
Db 78 ILDIGYDSSDKGFGDASCQGN-----IGIGAQSPDIAQGVDTAHETRVEGAADPLDAQGA 132

QY 154 ADRGEAAAAAVALDELVARF-LPM 174
Db 133 GDQGLMFGYAINDTPERMPLPI 154
```

Search completed: August 10, 2004, 15:42:28
Job time : 27.7312 secs

Blank

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:22:45 ; Search time 8.01937 Seconds
(without alignments)
2207.061 Million cell updates/sec

Title: US-09-455-978b-77
Perfect score: 933
Sequence: 1 MSNDNDTLVTADVRNGIDGH.....DELVARFLPMLKLLTFDQOI 184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	933	100.0	489	2	E84304	Htr10 transducer
2	933	100.0	489	2	T44978	transducer protein
3	148	15.9	439	2	A83713	methyl-accepting c
4	142	15.2	432	2	C69832	methyl-accepting c
5	132	14.1	537	2	C87302	methyl-accepting c
6	117.5	12.6	499	2	A37485	methyl-accepting c
7	117.5	12.6	499	2	A82703	methyl-accepting c
8	105.5	11.3	555	2	D87536	methyl-accepting c
9	99	10.6	448	2	A24436	IGG-binding protei
10	98	10.5	593	2	S00128	protein G precursor
11	97	10.4	218	2	T35174	hypothetical prote
12	94.5	10.1	883	2	A84210	hypothetical prote
13	94	10.1	1090	2	T00533	probable DNA2-NAM7
14	93.5	10.0	2048	2	C84609	hypothetical prote
15	93	10.0	881	2	AD2580	two component sens
16	93	10.0	881	2	C97362	protein sensor pro
17	92.5	9.9	503	2	AH3535	periplasmic dipept
18	89	9.5	955	2	A47334	Lckin kinesin-rela
19	88	9.4	500	2	B81060	N utilization subs
20	88	9.4	505	2	H81816	N utilization subs
21	88	9.4	775	2	T45238	probable transfera
22	87	9.3	469	2	B87094	probable molecular
23	86.5	9.3	157	2	F84349	hypothetical prote
24	85	9.1	2703	2	H81193	hemagglutinin/hemo
25	84.5	9.1	266	2	C84365	electron transfer
26	84	9.0	491	2	F96022	conserved hypothet
27	83.5	8.9	712	2	T02552	cellulose synthase
28	82.5	8.8	381	2	S28115	gas-vesicle protei
29	82	8.8	331	2	S01964	readthrough protei

ALIGNMENTS

RESULT 1

E84304

Htr10 transducer [imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001

C:Accession: E84304

R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.

Jung, K.H.; Alam, M.; Freitas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li

A>Title: Genome sequence of Halobacterium species NRC-1.

A:Reference number: A84160; MUID:20504483; PMID:11016950

A:Accession: E84304

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-489 <STO>

A:Cross-references: GB:AE004437; NID:gi10580997; PIDN:AAG19801.1; GSPDB:GN00138

C:Genetics:

C:Gene: htr10

C:Superfamily: Halobacterium salinarum transducer protein htr1

Query Match 100.0%; Score 933; DB 2; Length 489;
Best Local Similarity 100.0%; Pred. No. 1.9e-70;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60

Db 1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60

QY 61 ADALVTDFYDHLESYERTQDLFANSTKTVEQLKQAEYLLGLGRGEYDTEYAQRARIG 120

Db 61 ADALVTDFYDHLESYERTQDLFANSTKTVEQLKQAEYLLGLGRGEYDTEYAQRARIG 120

QY 121 KIHVLGLGPDVYLGAVTRYTGTLLDALDDVADRGEEAAAADVELVFLPMLKLITF 180

Db 121 KIHVLGLGPDVYLGAVTRYTGTLLDALDDVADRGEEAAAADVELVFLPMLKLITF 180

QY 181 DQOI 184

Db 181 DQOI 184

RESULT 2

T44978

transducer protein hemAT [validated] - Halobacterium salinarum

N:Alternate names: methyl-accepting taxis protein htB; transducer protein htB; transducer

C:Species: Halobacterium salinarum

C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 15-Sep-2000

C:Accession: T44978

R:Zhang, W.; Brooun, A.; McCandless, J.; Banda, P.; Alam, M.

Proc. Natl. Acad. Sci. U.S.A. 93, 4649-4654, 1996

30 82 8.8 372 2 G83074
31 82 8.8 750 1 D72530
32 82 8.8 911 2 AD2271
33 81.5 8.7 1893 2 T22661
34 81 8.7 302 2 H84184
35 81 8.7 1001 2 S74544
36 81 8.7 1311 2 T08986
37 81 8.7 2517 2 S58380
38 80.5 8.6 369 1 T44946
39 80.5 8.6 765 1 T44946
40 80 8.6 167 2 S52220
41 80 8.6 467 2 G82697
42 80 8.6 620 2 A83182
43 79.5 8.5 334 2 A95951
44 79.5 8.5 476 2 AE1139
45 79.5 8.5 547 2 F83720

glutamate 5-kinase
probable nicotine
cation-transportin
hypothetical prote
hypothetical prote
translation initia
hypothetical prote
probable RNA-direc
transducer protein
cobinamide kinase
hypothetical prote
hypothetical prote
hypothetical prote
Bacillus anthracis
2,4-diaminobutyrate

[illegible]

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolodziej, K.; Kuster, C.; Loms, D.; Maitav, T.; McInerney, J.M.; Miller, N.L.; Nelson, K.E.; Osofsky, J.S.; Paulsen, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapero, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A>Title: Complete Genome Sequence of *Caulobacter crescentus*.
A|Reference number: AB7249; MUID:21173698; PMID:11259647
A|Accession: C87302
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-537 <STO>
A|Cross-references: GB:AE005673; NID:gl3421593; PIDN:AAK22415.1; GSPDB:GNO0148
C|Genetics:
A|Gene: CC0428
C|Superfamily: methyl-accepting chemotaxis protein mcpA
Query Match 14.18; Score 132; DB 2; Length 537;

QY 33 IAWRLSFTGIDDDTAAALAEQPLFEATADALVTDYDHLSEYERTQDLFANSTKTVEQL 92
 Db 11 IGERTAFMGIDDKARSALRDLRPVIRAEIGKALDNFYGVKVRATPFRKFFSDRRHMAAS 70
 QY 93 KETQAEYLLGLGRGEYDTYAAQARIGIKHVDVLGPDVYLGAYTRYTGLLDALADDV 152
 Db 71 SRQQAHWGV-IAEGQSPDDYQAVRAIGQTHARIGLEPRWYGGYAVVGDHLVRAVDSM 129
 QY 153 -----VADRGEEAAAADVELVARELPM 174
 Db 130 WPRGLLAKGSDRAGEVAAL--MKAIFLDM 158
 RESULT 6
 A:Map position: circular chromosome
 A97485
 methyl-accepting chemotaxis protein mcpV (AF312877) [imported] - Agrobacterium tumefaciens
 C:Species: Agrobacterium tumefaciens
 C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
 C:Accession: A97485
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markels, B.;
 Science 294, 2323-2328, 2001
 A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A:Reference number: A97359; MUID:21608551; PMID:11743194
 A:Accession: A97485
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-499 <KUR>
 A:Cross-references: GB:AE007869; PIDN:AAK86834.1; PID:g15156046; GSPDB:GN00169
 C:Genetics:
 A:Gene: AGR_C1888
 A:Map position: circular chromosome
 Query Match 12.6%; Score 117.5; DB 2; Length 499;
 Best Local Similarity 26.5%; Pred. No. 0.034;
 Matches 35; Conservative 23; Mismatches 67; Indels 7; Gaps 3;
 QY 19 GHALADRIGLDEAEIAWRLSFTGIDDDTAAALAEQPLFEATADALVTDYDHLSEYERT 78
 Db 3 GOAKTDR-QLDE-----RLNFLGLGHERQNLSDMKGVITGSLDASLDRFYTKVRAVPET 56
 QY 79 QDLFANSTKTVEQLKETQAEYLLGLGRGEYDTYAAQARIGIKHVDVLGPDVYLGAYT 138
 Db 57 AKFF-SSEAHIIHAKSMQLKHSRIASGTFTNEDYNAVTAGRTHARLGLEPRWYGGYA 115
 QY 139 RYVTGLLDALAD 150
 Db 116 LMLDGIVKAVIE 127
 RESULT 7
 AB2703
 methyl-accepting chemotaxis protein mcpV [imported] - Agrobacterium tumefaciens (strain
 C:Species: Agrobacterium tumefaciens
 C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 C:Accession: AB2703
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; MUID:21608550; PMID:11743193
 A:Accession: AB2703
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-499 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AAAL42040.1; PID:g17739417; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: mcpV

A:Map position: circular chromosome

Query Match 12.6%; Score 117.5; DB 2; Length 499;
 Best Local Similarity 26.5%; Pred. No. 0.034;
 Matches 35; Conservative 23; Mismatches 67; Indels 7; Gaps 3;
 QY 19 GHALADRIGLDEAEIAWRLSFTGIDDDTAAALAEQPLFEATADALVTDYDHLSEYERT 78
 Db 3 GOAKTDR-QLDE-----RLNFLGLGHERQNLSDMKGVITGSLDASLDRFYTKVRAVPET 56
 QY 79 QDLFANSTKTVEQLKETQAEYLLGLGRGEYDTYAAQARIGIKHVDVLGPDVYLGAYT 138
 Db 57 AKFF-SSEAHIIHAKSMQLKHSRIASGTFTNEDYNAVTAGRTHARLGLEPRWYGGYA 115
 QY 139 RYVTGLLDALAD 150
 Db 116 LMLDGIVKAVIE 127

RESULT 8

D87536
 methyl-accepting chemotaxis protein McpM [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: D87536
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A>Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; MUID:21173698; PMID:11259647
 A:Accession: D87536
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-555 <STO>
 A:Cross-references: GB:AE005673; NID:gl3423838; PIDN:AAK24288.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC2317

Query Match 11.3%; Score 105.5; DB 2; Length 555;
 Best Local Similarity 24.7%; Pred. No. 0.39;
 Matches 40; Conservative 28; Mismatches 77; Indels 17; Gaps 5;
 QY 31 AEIAWRLSFTGIDDDTAAALAEQPLFEATADALVTDYDHLSEYERTQDLFANSTKTVE 90
 Db 46 AKLDORMAFMFRDERSRAHLFAIKFVIDAEIAGALGFYSQVFLFPDTRVKFRDDGH-MA 104
 QY 91 QLKETQAEYLLGLGRGEYDTYAAQARIGIKHVDVLGPDVYLGAYTRYTGLLDALAD 150
 Db 105 GAERAQAQAHWRRIAEAGYGSYVDVERIGRSHADADIAPOWYIGGYAVVVEVMRAL-- 162
 QY 151 DVVADRG-----EAAAADVELV-ARFLPM-LKLITF 180
 Db 163 --VAKRAKGLFNSAKSDAELADGLSALIKAAFLMDLSVSTY 202

RESULT 9

A24496
 Igg-binding protein - Streptococcus sp. (group G)
 C:Species: Streptococcus sp.
 C>Date: 17-Sep-1987 #sequence_revision 17-Sep-1987 #text_change 17-Mar-2000
 C:Accession: A24496; A39041
 R:Fameestock, S.R.; Alexander, P.; Nagle, J.; Filpula, D.
 J. Bacteriol. 167, 870-880, 1986
 A>Title: Gene for an immunoglobulin-binding protein from a group G Streptococcus.
 A:Reference number: A24496; MUID:86304178; PMID:3745123
 A:Accession: A24496
 A:Molecule type: DNA
 A:Residues: 1-448 <FAH>
 A:Cross-references: GB:M13825; NID:g153822; PIDN:AAA03664.1; PID:g153823
 R:Sjoebbring, U.; Bjoerck, L.; Kastern, W.
 J. Biol. Chem. 266, 399-405, 1991
 A>Title: Streptococcal protein G. Gene structure and protein binding properties.

A:Reference number: A39041; MUID:91093154; PMID:1985908
A:Accession: A39041
A>Status: preliminary
A:Molecule type: protein
A:Residues: 34-42, 'N', 45-48, 62-76, 186-200 <SUO>
C:Genetics:
A:Gene: spg
C:Superfamily: M5 protein
C:Keywords: transmembrane protein

Query Match 10.6%; Score 99; DB 2; Length 448;
Best Local Similarity 25.3%; Pred. No. 1;
Matches 48; Conservative 25; Mismatches 45; Indels 72; Gaps 8;
QY 10 TADVRNGIDGHADRIGLDEAEIAWRLSFTGIDD-----DTMAALAAE----- 53
Db 41 TPIIRNGGE--LTNLLGNSETTLARNEESATADLTAAAVADTVAAAAAENAGAAWEEA 97
QY 54 -----OPLFEATADAL-----VTFDYHLESYERTQDLFANSTKTVEOLKETQAEYLLG 102
Db 98 AAADALAKAKADALKFNKYGVSYYKNL-----INNKTVEGKDLQAOVW-- 145
QY 103 LRGEYDTEYAAQARIGIKIHVDVLGLGPDVYLGAIVRYVYTGLLDALADDVADRGEEAAA 162
Db 146 -----ESAKKARISEATD-----GLSDFLKSQTPA---EDTVK 175
QY 163 AVDELVARFL 172
Db 176 SIELAEAKVL 185

RESULT 10
S00128
protein G precursor - Streptococcus sp. (Streptococcus G148)
N:Alternate names: albumin-binding protein; cell wall-bound protein
C:Species: Streptococcus sp.
A:Variety: Streptococcus G148
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 19-May-2000
C:Accession: S00128; A27604; A26314
R:Olsson, A.; Eliasson, M.; Guss, B.; Nilsson, B.; Hellman, U.; Lindberg, M.; Uhlen, M.
Eur. J. Biochem. 168, 319-324, 1987
A:Title: Structure and evolution of the repetitive gene encoding streptococcal protein G
A:Reference number: S00128; MUID:88029445; PMID:3665928
A:Accession: S00128
A:Molecule type: DNA
A:Residues: 1-593 <OLS>
A:Cross-references: EMBL:X06173; NID:g47084; PIDN:CAA29540.1; PID:g47085
A>Note: the source is designated as Streptococcus G148
A:Note: part of this sequence, including the amino end of the mature protein, was confir
J. Sjoerding, U.; Falkenberg, C.; Nielsen, E.; Akerstroem, B.; Bjoerck, L.
J. Immunol. 140, 1595-1599, 1988
A:Title: Isolation and characterization of a 14-kDa albumin-binding fragment of streptoc
A:Reference number: A27604; MUID:88154455; PMID:2831269
A:Accession: A27604
A:Molecule type: protein
A:Residues: 62-101 <SUO>
R:Guss, B.; Eliasson, M.; Olsson, A.; Uhlen, M.; Frej, A.K.; Jorvall, H.; Flock, J.I.;
EMBO J. 5, 1567-1575, 1986
A:Title: Structure of the IgG-binding regions of streptococcal protein G.
A:Reference number: A26314; MUID:86300657; PMID:3017704
A:Accession: A26314
A:Molecule type: DNA
A:Residues: 114-593 <GUS>
A:Cross-references: GB:X04015; NID:g47071; PIDN:CAA27638.1; PID:g47072
C:Function:
A:Description: it is part of the cell wall structure of group G streptococci and is cova
C:Superfamily: M5 protein
C:Keywords: duplication; membrane protein
F:1-33/Domain: signal sequence #status predicted <SIG>
F:34-593/Product: protein G #status experimental <MAT>
F:34-116/Domain: alanine-rich <ALA>
F:117-290/Domain: AB duplication <DUPL>
F:117-140/Region: A repeat

F:141-191/Region: B
F:192-215/Region: A repeat
F:216-266/Region: B
F:267-290/Region: A repeat
F:303-497/Domain: IgG binding <IGB>
F:303-357/Region: C repeat
F:358-372/Region: D
F:373-427/Region: C repeat
F:428-442/Region: D
F:443-497/Region: C repeat
F:498-567/Domain: proline-rich <PRO>
F:568-593/Domain: carboxyl-terminal <CTD>

Query Match 10.5%; Score 98; DB 2; Length 593;
Best Local Similarity 25.3%; Pred. No. 1.8;
Matches 48; Conservative 25; Mismatches 45; Indels 72; Gaps 8;
QY 10 TADVRNGIDGHADRIGLDEAEIAWRLSFTGIDD-----DTMAALAAE----- 53
Db 41 TPIIRNGGE--LTNLLGNSETTLARNEESATADLTAAAVADTVAAAAAENAGAAWEEA 97
QY 54 -----OPLFEATADAL-----VTFDYHLESYERTQDLFANSTKTVEOLKETQAEYLLG 102
Db 98 AAADALAKAKADALKFNKYGVSYYKNL-----INNKTVEGKDLQAOVW-- 145
QY 103 LRGEYDTEYAAQARIGIKIHVDVLGLGPDVYLGAIVRYVYTGLLDALADDVADRGEEAAA 162
Db 146 -----ESAKKARISEATD-----GLSDFLKSQTPA---EDTVK 175
QY 163 AVDELVARFL 172
Db 176 SIELAEAKVL 185

RESULT 11
T35174
hypothetical protein SC5A7.25c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 04-Mar-2000
C:Accession: T35174
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1998
A:Reference number: Z21570
A:Accession: T35174
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-218 <SEE>
A:Cross-references: EMBL:AL031107; PIDN:CAA19954.1; GSPDB:GN00070; SCOEDB:SC5A7.25c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC5A7.25c
C:Superfamily: Streptomyces coelicolor hypothetical protein SC5A7.25c

Query Match 10.4%; Score 97; DB 2; Length 218;
Best Local Similarity 29.6%; Pred. No. 0.62;
Matches 53; Conservative 21; Mismatches 77; Indels 28; Gaps 10;
QY 7 TLVTADVNRGIDGHADRIGLDEAEI--AWRLS-FTGIDDDTMAALAAEQPLFEATADA 63
Db 15 TLAADVVRVG-----ALADRLDVPHAEVFDVGRLSAASGVPSVVGALLGRP-----TGGA 66
QY 64 LV-TDFYHLESYERTQDLFANSTKTVEQ-----LKETQAEYLLGLGRGEYDTEYAAQ 115
Db 67 DVQTRFVRQLDLLRLTR-LKPNGRKKTQEIADGAGMSRQQAGALIN-GDRRPTMEHCDA 124
QY 116 RARIGIKHD--VLGLGPDVYLGAIVRYVYTGLLDALADDVADRGEEAAAADVELVARFL 172
Db 125 IQRFFRVHAGFLTAEDPEALAGALQRTQELQQL-----ADREKQAAAPADPLRL 178

RESULT 12
A84210
hypothetical protein Vng0514c [imported] - Halobacterium sp. NRC-1

QY 55 PLFE---ATADALVTDFYDHUESYERTQDLFA-----NSTKTVEQLKETQAEYLLGL 103

A; Experimental source: strain C58 (Dupont)

C;Genetics:

A:Gene: Atu0027

A;Map position: circular chromosome

	Query Match	10.0%;	Score 93;	DB 2;	Length 881;
	Best Local Similarity	28.7%;	Pred. No. 7.7;		
	Matches	49;	Conservative	23;	Mismatches 57;
					Indels 42;
					Gaps 10;
QY	37	LSFTGDDDTMAALAEQPLFEAT----	ADALVTDFYDHLESYERTQDL-----	FANST	86
DB	621	LTFVNMTD----	SVRAERALKKNDALLKADELKNDVFQHV-SYELRSPNTNIIGTDL		675
QY	87	KT--VEQLKETQAEYLLGLGRGEYDT EYAQAQRARI	QKIHDVLGIGPDPVYLQ-AYTRYTGT		143
DB	676	KTPGIGQLTERQAEYL-----	DHISTSSSVLLTIVNDILDLA-TVDAGIMQLNYS	SDN	756
QY	144	LILDALADDV---	VADRGEEAAAADVELVARFLP-----	MUKLIT	179
DB	727	ELNELLDVSVQIADLSQSGISLEIVAPAHGLSVADHORLQILTKL		777	

Search completed: August 10, 2004, 15:29:28
Job time : 9.01937 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:19:00 ; Search time 5.79177 Seconds
(without alignments)
1654.229 Million cell updates/sec

Title: US-09-455-978b-77

Perfect score: 933

Sequence: 1 MSNDNDTLVTADVRNGIDGH.....DELVARFLPMLKLLTFDQOI 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	DB	ID	Description
1	933	100.0	489	1	HMAT_HALN1	Q9hpr6 halobacteri
2	933	100.0	489	1	HMAT_HALSA	P71413 halobacteri
3	142	15.2	432	1	HMAT_BACSU	O07621 bacillus su
4	99	10.6	448	1	SPG1_STRSG	P06654 streptococ
5	98	10.5	593	1	SPG2_STRSG	P19909 streptococ
6	90.5	9.7	305	1	LPXC_RALSO	Q9xhl6 ralstonia s
7	89	9.5	955	1	KINL_LEICH	P46865 leishmania
8	88	9.4	775	1	PLSB_MYCLE	Q9x7b0 mycobacteri
9	87	9.3	469	1	TIG_MYCLE	Q9cbv2 mycobacteri
10	85.5	9.2	957	1	SECA_MYCSM	P71533 mycobacteri
11	82.5	8.8	381	1	GVPC_HALME	O02228 halobacteri
12	82	8.8	331	1	VAL_BPSP	P09677 bacterioph
13	82	8.8	372	1	PROB_PSEAE	Q9hvl9 pseudomonas
14	81	8.7	1001	1	IF2_SYNY3	P72689 synecocyst
15	80.5	8.6	764	1	HTR2_HALSA	P71410 halobacteri
16	80.5	8.6	765	1	DDP4_BOVIN	P81425 bos taurus
17	80.5	8.6	765	1	DDP4_FELCA	Q9n2i7 felis silve
18	80.5	8.6	867	1	SVA_FUSNN	Q8rfj8 fusobacteri
19	80	8.6	461	1	KGPF_THEZI	Q9hnl1 thermococcu
20	80	8.6	954	1	KGPF_PSESM	Q98715 pseudomonas
21	79.5	8.5	334	1	YUHP3_RHIME	Q92v44 rhizobium m
22	79.5	8.5	763	1	HTR2_HALN1	Q9hpr1 halobacteri
23	79.5	8.5	1415	1	RPOC_HAEIN	P43739 haemophilus
24	78.5	8.4	374	1	PROB_RALSO	Q9xvll ralstonia s
25	78.5	8.4	382	1	GVCL_HALN1	P24574 halobacteri
26	78.5	8.4	454	1	KGPF_PYRPU	Q9v2v7 pyrococcus
27	78.5	8.4	584	1	Y742_STRCO	Q9znbo streptomyce
28	78.5	8.4	766	1	DDP4_HUMAN	P27487 homo sapien
29	78.5	8.4	949	1	SEAL_MYCTU	O5885 mycobacteri
30	78	8.4	442	1	HRB6_XANCV	P80153 xanthomonas
31	78	8.4	450	1	KGPF_PYRHO	O59355 pyrococcus
32	77.5	8.3	459	1	KGPF_THELI	Q977q3 thermococcu
33	77	8.3	323	1	T2B1_BAGST	P70985 bacillus st

34	77	8.3	372	1	PROB_PSESM	Q889f0 pseudomonas
35	77	8.3	571	1	V096_FOWPV	Q9j5c4 fowlpox vir
36	77	8.3	591	1	IF2F_METMA	Q8pu78 methanosarc
37	76.5	8.2	451	1	K6PF_PYRAB	Q9vla6 pyrococcus
38	76.5	8.2	483	1	SYC_SYNY3	P74330 synecocyst
39	76.5	8.2	495	1	ENGX_YERPE	Q8zct9 yersinia pe
40	76.5	8.2	753	1	CAT2_NEUCR	Q8xi82 neurospora
41	76.5	8.2	940	1	SECA_STRGR	P95759 streptomyc
42	76.5	8.2	1683	1	YJD9_YEAST	P47054 saccharomyc
43	76	8.1	171	1	RS13_HALN1	Q9v2w4 halobacteri
44	76	8.1	239	1	VG88_BPML5	Q05305 mycobacteri
45	76	8.1	654	1	SYT_LACPL	Q88w49 lactobacilli

ALIGNMENTS

RESULT 1

ID	HMAT_HALN1	STANDARD;	PRT;	489 AA.
AC	Q9HPR6;			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Heme-based aerotactic transducer hemAT.			
GN	HEMAT OR HTR10 OR VNG1505G.			
OS	Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).			
OC	Archaea; Euryarchaeota; Halobacteria; Halobacteriales;			
OC	Halobacteriaceae; Halobacterium.			
OX	NCBI_TaxID=64091;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20504483; PubMed=11016950;			
RA	Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,			
RA	Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,			
RA	Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,			
RA	Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,			
RA	Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,			
RA	Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,			
RA	Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,			
RA	Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.,			
RA	"Genome sequence of Halobacterium species NRC-1.";			
Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).				
CC	!- FUNCTION: Heme-containing signal transducer responsible for			
CC	aerotaxis, the migratory response toward or away from oxygen (By			
CC	similarity).			
CC	!- SUBUNIT: Homotrimer (Probable).			
CC	!- SIMILARITY: Contains 1 methyl-accepting transducer domain.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>) or send an email to license@isb-sib.ch.

EMBL; AE005064; AAG19801.1; -
DR PIR; E84304; E84304.
DR HSSP; P02942; 1QV7.
DR InterPro; IPR004089; Chmtaxis_trans.
DR InterPro; IPR004090; Me_chemotaxis.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
KW Transducer; Heme; Complete proteome.
FT DOMAIN 218 454 METHYL-ACCEPTING TRANSDUCER.
SQ SEQUENCE 489 AA; 52852 MW; 8C0AEF179667791E CRC64;

Query Match 100.0%; Score 933; DB 1; Length 489;
Best Local Similarity 100.0%; Pred. No. 4.1e-68;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

RESULT 3	HMAT_BACSU	STANDARD;	PRT;	432 AA.
ID	HMAT_BACSU			
AC	O07621;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Heme-based aerotactic transducer hemAT.			
GN	HMAT OR BSU10380.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98040224; PubMed=9579061;			
RA	Noback M.A., Holsappel S., Kiewiet R., Terpstra P., Wambutt R.,			
RA	Wedler H., Venema G., Bron S.;			
RT	"The 172 kb prkA-addAB region from 83 degrees to 97 degrees of the			
RT	Bacillus subtilis chromosome contains several dysfunctional genes,			
RT	the glyB marker, many genes encoding transporter proteins, and the			
RT	ubiquitous hit gene."			
RL	Microbiology 144:859-875(1998).			
[2]				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98044033; PubMed=9384377;			
RA	Kuznet F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,			
RA	Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,			
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,			
RA	Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,			
RA	Ghm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,			
RA	Guiseppi G., Guy B.J., Haga K., Hatech J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.			
RA	Joris B., Karamata D., Kaehara Y., Klaerr-Blanchard M., Klein C.,			
RA	Kobayashi Y., Kottner P., Koningsstein G., Krogh S., Kumano M.,			
RA	Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,			
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,			
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,			
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,			
RA	Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,			
RA	Presecan E., Pujic P., Purnelle B., Rapoport G., Rev M., Reynolds			
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,			
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,			

RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*.";
RL Nature 390:249-256 (1997).
RN [3]
RP FUNCTION
RX MEDLINE=20140131; PubMed=10676961;
RA Kou S., Larsen R.W., Boudko D., Riley C.W., Karatan E., Zimmer M.,
RA Ordal G.W., Alam M.;
RT "Myoglobin-like aerotaxis transducers in *Archaea* and *Bacteria*.";
RL Nature 403:540-544 (2000).
RN [4]
RP RESONANCE RAMAN SPECTROSCOPY, AND SUBUNIT.
RX MEDLINE=21950695; PubMed=11821422;
RA Aono S., Kato T., Matsuki M., Nakajima H., Ohta T., Uchida T.,
RA Kitagawa T.;
RT "Resonance Raman and ligand binding studies of the oxygen-sensing
RT signal transducer protein HemaR from *Bacillus subtilis*.";
RL J. Biol. Chem. 277:13528-13538 (2002).
CC -!- FUNCTION: Heme-containing signal transducer responsible for
CC aerotaxis, the migratory response toward or away from oxygen.
CC -!- SUBUNIT: Homotetramer (Probable).
CC -!- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL; Y14084; CAA74545.1; -;
DR EMBL; Z99109; CAB12878.1; -;
DR PIR; C69832; C69832.
DR Subtilist: BG13066; hemaR.
DR GO; GO:0020037; F:heme binding; IDA.
DR GO; GO:0004871; F:signal transducer activity; IDA.
DR GO; GO:0009454; P:aerotaxis; IDA.
DR GO; GO:0007165; P:signal transduction; IDA.
DR InterPro; IPR004089; Cmtaxis_transd.
DR Pfam; PF00015; MCFsignal; 1.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS0111; CHEMOTAXIS_TRANSDUC_2; 1.
KW Transducer; Heme; Complete proteome.
FT DOMAIN 184 420 METHYL-ACCEPTING TRANSDUCER.
SQ SEQUENCE 432 AA; 48767 MW; 87D0655974070B64 CRC64;
Query Match 15.2%; Score 142; DB 1; Length 432;
Best Local Similarity 25.3%; Pred. No. 0.00027;
Matches 39; Conservative 29; Mismatches 76; Indels 10; Gaps 2;
Qy 31 AEIAWRLSFTGIDDDTMAALAAEOPLEATDALVTDYDHLSEYRTQDLPANSTKTVE 90
Db 32 ADVKQKLWVGLGDAELVLEQLQFLQENIVNIDAFYKNDHSSLLMDII-NDHSSVD 90
Qy 91 QLKETOAYLLGLGRGEYDTYAAQARAIKGIHVDVGLGPDVYLGAIVRYTGLLDALAD 150
Db 91 RLKQTLKRIHQEMFAGVIDDDFEIEKRNRIASIHRIKGLLPKRWYGAFOELLMSIDYI- 148
Qy 151 DVVADGEEAAAVDVARLPLMLKLLTFDQOI 184
Db 149 -----EASITNQQLLKAIKATTKIINLEQQL 175
RESULT 4
SPGI_STRSG

ID AC SPGI_STRSG STANDARD; PRT; 448 AA.
DT P06654;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Immunoglobulin G binding protein G precursor (IGG binding protein G).
GN SPG.
OS Streptococcus sp. (Lancefield group G).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1320;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86304178; PubMed=3745123;
RA Rahnestock S.R., Alexander P., Nagle J., Filpula D.;
RT "Gene for an immunoglobulin-binding protein from a group G
RT streptococcus.";
RL J. Bacteriol. 167:870-880 (1986).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.07 ANGSTROMS) OF 228-282.
RX MEDLINE=94213848; PubMed=8161530;
RA Gallagher T., Alexander P., Bryan P., Gilliland G.L.;
RT "Two crystal structures of the B1 immunoglobulin-binding domain of
RT streptococcal protein G and comparison with NMR.";
RL Biochemistry 33:4721-4729 (1994).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 293-351.
RX MEDLINE=95055731; PubMed=7966308;
RA Derrick J.P., Wigley D.B.;
RT "The third IGG-binding domain from streptococcal protein G. An
RT analysis by X-ray crystallography of the structure alone and in a
RL complex with Fab.";
RN [4]
RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 297-352.
RX MEDLINE=95308043; PubMed=7788293;
RA Sauer-Briksson A.E., Kleywegt G.J., Uhlen M., Jones T.A.;
RT "Crystal structure of the C2 fragment of streptococcal protein G in
RL complex with the Fc domain of human IgG.";
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 228-282.
RA Butterworth S., Lamzin V.S., Wigley D.B., Derrick J.P., Wilson K.S.;
RL Submitted (APR-1997) to the PDB data bank.
RN [6]
RP STRUCTURE BY NMR OF 298-351.
RX MEDLINE=91335209; PubMed=1871600;
RA Gronenborn A.M., Filpula D.R., Essig N.Z., Achari A., Whitlow M.,
RA Wingfield P.T., Clore G.M.;
RT "A novel, highly stable fold of the immunoglobulin binding domain of
RT streptococcal protein G.";
RL Science 253:657-661 (1991).
CC -!- FUNCTION: BINDS TO THE CONSTANT FC REGION OF IGG WITH HIGH
CC AFFINITY.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL; M13825; AAA03664.1; -;
DR PIR; A24496; A24496.
DR PDB; 1EM7; 08-MAY-02.
DR PDB; 1FCC; 20-JUL-95.
DR PDB; 1GBI; 15-APR-93.
DR PDB; 1IGG; 03-JUN-95.
DR PDB; 1IGD; 01-NOV-94.
DR PDB; 1MPE; 30-OCT-02.

DR PDB; 1MVK; 30-OCT-02.
DR PDB; 1PGA; 30-APR-94.
DR PDB; 1PGB; 30-APR-94.
DR PDB; 1PGX; 15-JUL-92.
DR PDB; 2GBL; 15-APR-93.
DR PDB; 2IGD; 29-JUL-98.
DR PDB; 2IGH; 31-JAN-94.
DR InterPro; IPR002988; GA.
DR InterPro; IPR005877; Gpos_Ysirk.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000724; IgG_bind_B.
DR Pfam; PF01468; GA; 2.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF01378; IgG_binding_B; 2.
DR Pfam; PF04650; Ysirk_signal; 1.
DR PRINTS; PR00015; GPOSANCHOR.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMs; TIGR01168; Ysirk_signal; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
KW IgG-binding protein; Cell wall; Peptidoglycan-anchor; Repeat; Signal;
KW 3D-structure.
FT SIGNAL 1 33
FT CHAIN 34 417 IMMUNOGLOBULIN G BINDING PROTEIN G.
FT PROPEP 418 448 REMOVED BY SORTASE (POTENTIAL).
FT DOMAIN 69 111 ALA-RICH.
FT DOMAIN 104 215 2 X 37 AA REPEATS.
FT REPEAT 104 140 1-1.
FT REPEAT 179 215 1-2.
FT DOMAIN 228 352 2 X 55 AA REPEATS.
FT REPEAT 228 282 2-1.
FT REPEAT 238 352 2-2.
FT DOMAIN 386 410 5 X 5 AA REPEATS OF [DE]-D-A-K-K.
FT SITE 414 418 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 417 417 AMIDE-LINKED TO CELL WALL (POTENTIAL).
FT STRAND 228 234
FT STRAND 239 245
FT HELIX 249 262
FT TURN 263 264
FT STRAND 268 272
FT TURN 273 276
FT STRAND 277 281
FT STRAND 309 316
FT STRAND 319 332
FT TURN 333 334
FT STRAND 338 342
FT TURN 343 346
FT STRAND 347 351
SQ SEQUENCE 448 AA; 47567 MW; A0759060C8F9E6CA CRC64;
Query Match 10.6%; Score 99; DB 1; Length 448;
Best Local Similarity 25.3%; Pred. No. 0.84;
Matches 48; Conservative 25; Mismatches 45; Indels 72; Gaps 8;
QY 10 TADVRNGIDCHALADRIGLDEAEIANRLSFTGIDD-----DTMAALAAE----- 53
DB 41 TPIIRNGGE---LTNLLGNSETTLALRNEESATADLTAAAVADTVAAAPAAENAGAAWEA 97
QY 54 ----QLFEATADAL-----VTDFYDHLESYERTQDLFPANSTKTVEQKETAQAEYLLG 102
DB 98 AAAADALAKAKADALKEFNKYGVSDYKYL-----INNKTVEGKDLQAQVW-- 145
QY 103 LGRGEYDTEVAQAQARKTGKIHVLGLGPVILGAYTRYTYTGLLDALADDDVADRGEEAAA 162
DB 146 -----ESAKKARISEATD-----GLSDFLKSQTPA---EDTVK 175
QY 163 AVELVARFL 172
DB 176 SIELAEAKVL 185
RESULT 5
SPG2_STRSG


```

OX NCBI_TaxID=44271;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MHOM/BR/82 / Isolate BA-2;
RA MEDLINE=93133867; PubMed=8421715;
RX Burns J.M. Jr., Shreffler W.G., Benson D.R., Ghalib H.W., Badaro R.,
RA Read S.G.;
RT "Molecular characterization of a kinsin-related antigen of
RT Leishmania chagasi that detects specific antibody in African and
RA American visceral leishmaniasis."
RL Proc. Natl. Acad. Sci. U.S.A. 90:775-779(1993).
CC -!- DEVELOPMENTAL STAGE: PREDOMINANT IN AMASTIGOTES.
CC -!- SIMILARITY: Belongs to the kinsin-like protein family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L07879; AAA29254.1; -
CC PIR; A47334; A47334.
CC HSPG; P17119; 3KAR.
CC InterPro: IPR001752; kinsin_motor.
CC Pfam; PF00225; kinsin; 1.
CC PRINTS; PR00380; KINESINHEAVY.
CC SMART; SM00129; KISC; 1.
CC PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
CC PROSITE; PS0067; KINESIN_MOTOR_DOMAIN2; 1.
CC Motor protein; Microtubule; ATP-binding; Coiled coil; Repeat.
KW DOMAIN 1 399 KINESIN_MOTOR (BY SIMILARITY).
FT DOMAIN 1 399
FT NP_BIND 426 >955 COILED COIL (POTENTIAL).
FT NP_BIND 122 129 ATP (POTENTIAL).
FT DOMAIN 704 >955 7 X 39 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 704 742 1.
FT REPEAT 743 781 2.
FT REPEAT 782 820 3.
FT REPEAT 821 859 4.
FT REPEAT 860 898 5.
FT REPEAT 899 937 6.
FT REPEAT 938 >955 7 (PARTIAL).
FT NON_TER 955 955
SQ SEQUENCE 955 AA; 106168 MW; 8CA76815BE84C6E9 CRC64;

Query Match 9.5%; Score 89; DB 1; Length 955;
Best Local Similarity 29.3%; Pred. No. 12;
Matches 49; Conservative 19; Mismatches 69; Indels 30; Gaps 7;

QY 44 DDTMAALAAEQPLFEATADALVT---DFYDHLSEYERTQDLFANSTKVEQLKETQAEVL 100
DB 546 ESTVAQLREQREREVALDQTHQKQLEAESSERTA---AERDQLLQQLTELQSE-R 601
QY 101 LGLGRGEYDTE-YAAQRAIRIGIHVDVLGLGPDVYLGA-----YTRYTIGLL----- 145
DB 602 TQLSQVTDRELTLDLQIQEYGETELARDVALCAQEMEARVHAAPVFLQLLELAT 661
QY 146 ---DALADDVADRGEAAAADVLV-----ARFLPMKLKLFDDQOI 184
DB 662 EWEDALRRALAEERDEAAAEALDAASTQNAESACERLTSLQQL 708

RESULT 8
ID PLSB MYCLE STANDARD; PRT; 775 AA.
AC Q9X7B0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glycero-3-phosphate acyltransferase (EC 2.3.1.15) (GPAT).
GN PLSB OR ML1246 OR MLCB1610.07.
OS Mycobacterium leprae.

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1769;
SEQUENCE FROM N.A.
STRAIN=TN;
MEDLINE=21128732; PubMed=11234002;
Cole S.T., Eigleier K., Parkhill J., James K.D., Thomson N.R.,
Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R.M., Devlin K., Duthoy S., Feltingwell T., Fraser A., Hamlin N.,
Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
Barrell B.G.;
"Massive gene decay in the leprosy bacillus."
Nature 409:1007-1011(2001).
-!- CATALYTIC ACTIVITY: Acyl-CoA + sn-glycerol 3-phosphate = CoA + 1-
acyl-sn-glycerol 3-phosphate.
-!- PATHWAY: De novo phospholipid biosynthesis; first step. May also
function in the regulation of membrane biogenesis.
-!- SUBCELLULAR LOCATION: Membrane-bound (By similarity).
-!- SIMILARITY: Belongs to the GPAT / DAPAT family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL049913; CAB43153.1; -
CC EMBL; AL583921; CAC31627.1; -
CC PIR; T45238; T45238.
CC Leproma; ML1246; -.
CC HAMAP; MF_00393; -; 1.
CC InterPro: IPR002123; Acyltransferase.
CC Pfam; PF01553; Acyltransferase; 1.
CC SMART; SM00563; PISC; 1.
CC Phospholipid biosynthesis; Transferase; Acyltransferase; Membrane;
Complete proteome.
KW COMPLETE PROTEOME.
SQ SEQUENCE 775 AA; 87363 MW; 03DD77C778293CDF CRC64;

Query Match 9.4%; Score 88; DB 1; Length 775;
Best Local Similarity 29.4%; Pred. No. 12;
Matches 40; Conservative 20; Mismatches 40; Indels 36; Gaps 7;

QY 32 ETAWRLSF---TGIDDDTMAALAAEQPLFEATADALVTDFYDHLSEYERTQDLFANSTK 87
DB 631 ETAWHQNEDRVSGGDGDDIDAMLLTKRPLI---SDAMLRVFF---EAYDIVADVLRDAPA 684
QY 98 TVEQKETQAEYLLGLGRGEYDTEYAAQ-RARIGIKHVDVLGLGPDVYLGAIFYTRYTGILLD 146
DB 685 DVGQKELT-ELALGVGR-----QYVAQGRVRSGE-----SVSTLLF 719
QY 147 ALADDVADRGEAAA 162
DB 720 ATAYQVVVDQNLIAFA 735

RESULT 9
ID TIG MYCLE STANDARD; PRT; 469 AA.
AC Q9CBY2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trigger factor (TF).
GN TIG OR ML1481.
OS Mycobacterium leprae.
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
```



```

DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Gas vesicle protein C.
GN GVPC.
OC Halobacterium mediterranei (Haloferax mediterranei).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Haloferax.
OX NCBI_TaxID=2252;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1411;
RX MEDLINE=93021102; PubMed=1404376;
RA Englert C., Krueger K., Offner S., Pfeifer F.;
RT "Three different but related gene clusters encoding gas vesicles in
RL halophilic archaea.";
RL J. Mol. Biol. 227:586-592(1992).
CC CC -I- FUNCTION: May confer stability to the gas vesicle membranes. Gas
CC vesicles are small, hollow, gas filled protein structures that are
CC found in several microbial planktonic microorganisms. They allow
CC the positioning of the organism at the favorable depth for growth.
CC CC -I- SUBCELLULAR LOCATION: Binds to the external surface of the gas
CC vesicle membrane.
CC CC -I- SIMILARITY: VERY DISTANT, TO GAS VESICLE PROTEINS TYPE C OF
CC CYANOBACTERIA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64701; CAA45944.1; -.
DR PIR; S28115; S28115.
DR InterPro; IPR008639; Halo GVPC.
DR Pfam; PF05465; Halo_GVPC; 1.
KW Gas vesicle; Repeat.
FT DOMAIN 22 274 7 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 22 59 1.
FT REPEAT 60 84 2.
FT REPEAT 85 122 3.
FT REPEAT 123 160 4.
FT REPEAT 161 192 5.
FT REPEAT 193 232 6.
FT REPEAT 233 274 7.
FT DOMAIN 246 335 ASP/GLU-RICH (ACIDIC).
FT SEQUENCE 381 AA; 42653 MW; 9FB4B199D0305921 CRC64;
Query Match 8.8%; Score 82.5; DB 1; Length 381;
Best Local Similarity 25.9%; Pred. No. 15;
Matches 45; Conservative 27; Mismatches 73; Indels 29; Gaps 9;
Qy 11 ADVENGIDGHALADRIGLDEAEIAWLSFTGIDDDTMAALAAEQPLFEATDALVTFYD 70
Db 81 ADREDAFD--AYADIFATDVAEMQDVSLAAIIDLRAEMDETTEAFDAYADAFVTD--- 135
Qy 71 HLESYERTQDLFANSTKTVQELKQTAEYLLGLGR-GEYDEYAAQARIGKIHDLGLG 129
Db 136 -VATLRDVSU-----LTAISELQSEFVSQVEFGYSEFGAD---IDQFHAVVAEK 184
Qy 130 PDVY---LGAATRY---YTGIL-LDALADDVVVADRG-----EAAAAVDELVARF 171
Db 185 RDGHKQVADAPLQVREPHGVQSLDNTIAAFQREMGMGYRKAFETTEAPASF 238
RESULT 12
VAL_BPSP
ID -VAL_BPSP
AC P09677;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 15:22:05 ; Search time 22.0533 Seconds
(without alignments)
2632.505 Million cell updates/sec

Title: US-09-455-978B-77

Perfect score: 933

Sequence: 1 MSNDNDTLVTADVRNGIDGH.....DELVARFLPKLLTFDQOI 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	148	15.9	439	Q9KFH4	Q9kf4 bacillus ha
2	132	14.1	537	Q9AB06	Q9ab06 caulobacter
3	117.5	12.6	499	Q8UGU0	Q8ugl0 agrobacteri
4	105.5	11.3	555	Q9A5Y0	Q9a5y0 caulobacter
5	102	10.9	434	Q814K6	Q814k6 bacillus ce
6	98.5	10.6	1222	Q8D4B1	Q8d4b1 vibrio vuln
7	97.5	10.5	248	Q8S4W2	Q8s4w2 phytophthor
8	97	10.4	218	Q88031	Q88031 streptomyce
9	95.5	10.2	1364	Q84GK0	Q84gk0 escherichia
10	95	10.2	439	Q81JN0	Q81jn0 bacillus an
11	94.5	10.1	419	Q9W6U7	Q9w6u7 bacterioph
12	94.5	10.1	883	Q9HRW3	Q9hrw3 halobacteri
13	94	10.1	218	Q82MB8	Q82mb8 streptomyce
14	94	10.1	1090	Q64476	Q64476 arabidopsis
15	93.5	10.0	514	Q8FUX2	Q8fux2 brucella su
16	93	10.0	881	Q8UJA1	Q8uja1 agrobacteri

17	92.5	9.9	503	16	Q8YDG6	Q8ydg6 brucella me
18	92.5	9.9	514	2	Q8VOK3	Q8vok3 brucella ab
19	90.5	9.7	194	16	Q8DHH0	Q8dhh0 synchococc
20	89.5	9.6	400	16	Q82AV2	Q82av2 streptomyce
21	88.5	9.5	458	16	Q8PMR9	Q8pmr9 xanthomonas
22	88.5	9.5	505	11	Q9KK9	Q9kk9 mus musculu
23	88.5	9.5	1361	16	Q8RHV1	Q8rhv1 fusobacteri
24	88	9.4	500	16	Q9UYD3	Q9uyd3 neisseria m
25	88	9.4	505	16	Q9UTB6	Q9utb6 neisseria m
26	87.5	9.4	2246	2	Q9AKS6	Q9aks6 pseudomonas
27	86.5	9.3	157	17	Q9HNO8	Q9hnq8 halobacteri
28	86.5	9.3	468	16	Q8P6N2	Q8p6n2 xanthomonas
29	86.5	9.3	895	11	Q8CL17	Q8cl17 mus musculu
30	86	9.2	670	4	Q96CNS	Q96cns homo sapien
31	86	9.2	508	16	Q8CJH8	Q8cjk8 streptomyce
32	85.5	9.2	965	16	Q8IYS7	Q8iys7 bacillus an
33	85	9.1	306	16	Q8D7S6	Q8d7s6 vibrio vuln
34	85	9.1	2703	16	Q9KOT0	Q9koto neisseria m
35	84.5	9.1	266	17	Q9HND3	Q9hnd3 halobacteri
36	84.5	9.1	759	16	Q98C20	Q98c20 rhizobium l
37	84	9.0	170	3	Q87LH7	Q87lh7 neurospora
38	84	9.0	412	16	Q7UG88	Q7ug88 rhodospirell
39	84	9.0	421	2	Q9RQ25	Q9rq25 amycolatops
40	84	9.0	467	16	Q87DW5	Q87dw5 xylella fas
41	84	9.0	491	16	Q92TQ4	Q92tq4 rhizobium m
42	83.5	8.9	258	16	Q82G05	Q82gu5 streptomyce
43	83.5	8.9	712	10	Q80890	Q80890 arabidopsis
44	83.5	8.9	755	10	Q8RX83	Q8rx83 arabidopsis
45	83.5	8.9	826	16	Q92TC2	Q92tc2 rhizobium m

ALIGNMENTS

RESULT 1

Q9KFH4 PRELIMINARY; PRT; 439 AA.
AC Q9KFH4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Methyl-accepting chemotaxis protein.
GN BH0505.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331 (2000).
DR EMBL; AP001508; BAB04224.1; -.
DR PIR; A83713; A83713.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0005351; F:sugar porter activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. .; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis transd.
DR InterPro; IPR002114; Hbr SerpS.
DR Pfam; PF00015; MCPsignal; 1.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS01111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS00589; PTS_HPR_SER; 1.
KW Complete proteome.
SQ SEQUENCE 439 AA; 48918 MW; 1E092BB2F42592AE CRC64;


```

Db 3 GOAKTR-OLDE-----RLNFLGLGHERQNLSDMKGVITGSLDASLDRTYTKVRVPET 56
QY 79 QDLFANSTKTEQLKQAEYLLGLGRGEYDTYAAQARIGIKHVDVLGDPDVLGYAYT 138
Db 57 AKFF-SSEAHIIHAKSMQLKHSRIASGTFTNEDYTAIGRTHARLGLRPRWYIGYA 115
QY 139 RYTGLLDALAD 150
Db 116 LMLDGIVKAVIE 127

```

RESULT 4

```

Q9A5Y0 ID Q9A5Y0 PRELIMINARY; PRT; 555 AA.
AC Q9A5Y0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Methyl-accepting chemotaxis protein McpM.
GN CC2317.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C.; Feldblyum T.V.; Laub M.T.; Paulsen I.T.; Nelson K.E.,
RA Eisen J.; Heidelberg J.F.; Alley M.R.K.; Ohta N.; Maddock J.R.,
RA Potocka I.; Nelson W.C.; Newton A.; Stephens C.; Phadke N.D.; Ely B.,
RA DeBoy R.T.; Dodson R.J.; Durkin A.S.; Gwinn M.L.; Haft D.H.,
RA Kolonay J.F.; Smit J.; Craven M.B.; Khouri H.; Shetty J.; Berry K.,
RA Utterback T.; Tran K.; Wolf A.; Vanathavan J.; Ermolaeva M.; White O.,
RA Salzberg S.L.; Venter J.C.; Shapiro L.; Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005901; AA24288.1; -.
DR PIR; D87536; D87536.
DR HSSP; P02942; LQUT.
DR TIGR; CC2317; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chntaxis_transd.
DR Pfam; PF000360; HAMP.
DR SMART; SM00015; MCPsignal; 1.
DR SMART; SM00304; HAMP; 1.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS50885; HAMP; 1.
KW Complete proteome.
SQ SEQUENCE 555 AA; 59114 MW; C9560265C1EC0B31 CRC64;

```

```

Query Match 11.3%; Score 105.5; DB 16; Length 555;
Best Local Similarity 24.7%; Pred. No. 1.2;
Matches 40; Conservative 28; Mismatches 77; Indels 17; Gaps 5;

QY 31 AEIAWRLSFTGIDDDTMAALAAEQPLFEATADALVDFYDHLSEYRTQDLFANSTKIVE 90
Db 46 AKLDQMAFMFDRSRAHLRAIKPVIDAIEGAALGFYSQVRLFPDTRVKFRDDGH-MA 104
QY 91 QLKETOAEYLLGLGRGEYDTYAAQARIGIKHVDVLGDPDVLGYAYTRYTGLLDALAD 150
Db 105 GAERQAQAHWRRIAEAGGESYVRDVERIGRSHADADAPQWYIGGYAVVVEVMRAL-- 162
QY 151 DWADRG-----EAAAAVDLV-ARFLPM-LKLLTF 180
Db 163 --VAKRAGLFNSAKSDELADGLSALIKAAFLMDLSVSTY 202

```

RESULT 5

```

Q814K6 ID Q814K6 PRELIMINARY; PRT; 434 AA.
AC Q814K6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Methyl-accepting chemotaxis protein.
GN BC5424.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N.; Sorokin A.; Anderson I.; Galleron N.; Candelon B.,
RA Kapral V.; Bhattacharyya A.; Reznik G.; Mikhailova N.; Lapidus A.,
RA Chu L.; Mazur M.; Goltsman E.; Larsen N.; D'Souza M.; Walunas T.,
RA Grechkin Y.; Pusch G.; Haselkorn R.; Fonstein M.; Ehrlich S.D.,
RA Overbeek R.; Kyrpides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis."
RL Nature 423:87-91(2003).
DR EMBL; AE017015; AAP12286.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chntaxis_transd.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
KW Complete proteome.
SQ SEQUENCE 434 AA; 49441 MW; E1007BA0230620C5 CRC64;

```

```

Query Match 10.9%; Score 102; DB 16; Length 434;
Best Local Similarity 20.5%; Pred. No. 1.7;
Matches 34; Conservative 27; Mismatches 71; Indels 34; Gaps 3;

QY 31 AEIAWRLSFTGIDDDTMAALAAEQPLFEATADALVDFYDHLSEYRTQDLFANSTK--- 87
Db 34 SELKIQMMLHLSKEDLVVVKVLPFIYAEIDWITEKFY-----ANITQPN 80
QY 88 -----TVEQLKETOAEYLLGLGRGEYDTYAAQARIGIKHVDVLGDPDVLGYAYT 138
Db 81 LITTIERYSSIPKLKQTLTKIKELFSGNMHEDFTEQVRIAKRVQIGLHKKWYTAAYQ 140
QY 139 RYTGLLDALADVVADRGEAAAVDELVARFLPMLKLLTFDQOI 184
Db 141 ELFSIIKILOTKI-----TTIDDFSYSINVKLFTLEQEL 177

```

RESULT 6

```

Q8D4B1 ID Q8D4B1 PRELIMINARY; PRT; 1222 AA.
AC Q8D4B1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Conserved hypothetical protein.
GN VV21397.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H.; Kim S.Y.; Chung S.S.; Kim J.J.; Moon Y.H.; Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6."

```

RL	Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AE016813; AAC08278.1; -	
DR	InterPro; IPR000437; Prok lipoprot S.	
DR	PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.	
KW	Hypothetical protein; Complete proteome.	
SQ	SEQUENCE 1222 AA; 136974 MW; B04525D301FA99E2 CRC64;	
	Query Match 10.6%; Score 98.5; DB 16; Length 1222;	
	Best Local Similarity 20.2%; Pred. No. 15;	
	Matches 37; Conservative 34; Mismatches 59; Indels 53; Gaps 7;	
QY	43 DDDTMAALAAEPLFEATADALVTDFYHLESY-----ERTODLFANSTKV 89	
DB	606 DSENTYRLTEHQPFIWSSPTDSSYEBYADY--NYNQKWNENTAYVWVTELLADSSIKV 563	
QY	90 -----EQKKEQAE-----YLLGLGRGEYDT-----EYAAQ 115	
DB	664 SGRVYLLADSDAIKETSSKPGKPFSEFSLHRTISATEKRAYETWHHIALEGYDFTAS 723	
QY	116 RARIQKIHVDVLGLGPDVYL-GAYTRYTGLLDALADDVVADRGEEAAAADVELVARFLPM 174	
DB	724 ANVFGHQYTVFNLYDIIWGYSTQWFDNRNIDLIAE---VDRLLQGQVFPDQITPEMIEN 780	
QY	175 LKL 177	
DB	781 LNL 783	
	RESULT 7	
ID	Q8S4W2 PRELIMINARY; PRT; 248 AA.	
AC	Q8S4W2;	
DT	01-JUN-2002 (TrEMBLrel. 21, Created)	
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE	Syntaxin 6.	
OS	Phytophthora infestans (Potato late blight fungus).	
OC	Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;	
OC	Phytophthora.	
OX	NCBI_TaxID=4787;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=2194782; PubMed=11950882;	
RA	Dacks J.B., Doolittle W.F.;	
RT	"Novel syntaxin gene sequences from Giardia, Trypanosoma and algae:	
RT	implications for the ancient evolution of the eukaryotic endomembrane	
RT	system.";	
RL	J. Cell Sci. 115:1635-1642(2002).	
DR	EMBL; AF404749; AAM12665.1; -	
DR	InterPro; IPR000727; T_SNARE.	
DR	Pfam; PF05739; SNARE; 1.	
DR	SMART; SM00397; T_SNARE; 1.	
DR	PROSITE; PS50192; T_SNARE; 1.	
SQ	SEQUENCE 248 AA; 27934 MW; 2FE6891BA5AA0067 CRC64;	
	Query Match 10.5%; Score 97.5; DB 10; Length 248;	
	Best Local Similarity 22.7%; Pred. No. 1.9;	
	Matches 42; Conservative 31; Mismatches 67; Indels 45; Gaps 7;	
QY	6 DTLVTADVNRGIDGHALADRIGLDEAEIAWRLSF-----TGIDDDDTMAA--LA 51	
DB	68 ETIIVWEANRAKEH-----IDAEIASRKAFVAATRKELQAVSTEISTDIVKTRIK 120	
QY	52 AEQPLFEATADALVTDFYHLESYERTQDLFANSTKVTEQKTOAEYLLGLGRGEYDTE 111	
DB	121 EERKLMPAKSS--TSFRNLTGQERNERFLEDETQROQQIMQFONDSLGL----- 170	
QY	112 YAAQARAIKIHVDVLGLGPDVYL-GAYTRYTGLLDALADDVVADRGEEAAAADVELVARF 171	
DB	171 ----HSDITRLH-----GVTVESISSEVKQNKMLDLDLTDVV-----DEAQERWNVWGRL 216	
QY	172 LPLMK 176	
	Query Match 10.4%; Score 97; DB 16; Length 218;	
	Best Local Similarity 29.6%; Pred. No. 1.7;	
	Matches 53; Conservative 21; Mismatches 77; Indels 28; Gaps 10;	
QY	7 TLVTADVNRGIDGHALADRIGLDEAEI--AWRLS-FTGIDDDTMAALAAEQPLFEATADA 63	
DB	15 TLAADVAVRG-----ALADRLDVPHEVDFVGRUSAAAGSPESVVGALLGRP---TGGA 66	
QY	64 LV-TDFVDHLESYERTQDLFANSTKVTEQ-----LKETOAEYLLGLGRGEYDTEYAAQ 115	
DB	67 DVQTRFVQRDLRLRRTE-LKPNGRKYTQOEIADGAGMSRQAGALIN-GDRRPTMERHCA 124	
QY	116 RARIQKIH--VLGLGPDVYL-GAYTRYTGLLDALADDVVADRGEEAAAADVELVARFL 172	


```

Db 125 IQRFFRVHAGFLTAEDPEALAGALQTEQLLQQL-----ADRERQAAAPADDDPLERLL 178

RESULT 9
Q84GKO PRELIMINARY; PRT; 1364 AA.
AC Q84GKO;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Secreted autotransporter protein Bata.
GN BATA.
OS Escherichia coli.
OG Plasmid pCS1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H10407;
RA Fleckenstein J.M., Patel S.K., Dotson J.;
RT "Identification and molecular characterization of Bata, an
RT autotransporter protein of enterotoxigenic Escherichia coli.";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY163491; AA017297.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0044252; F:serine-type endopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR000710; Peptidase S6.
DR InterPro; IPR004899; Pertactin.
DR Pfam; PF02395; IGAL1; 1.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR00921; IGASERPTASE.
DR TIGRfams; TIGR01414; autotrans_barl; 1.
KW Plasmid.
SQ SEQUENCE 1364 AA; 147696 MW; 33B341FDAB6859E2 CRC64;

Query Match 10.2%; Score 95.5; DB 2; Length 1364;
Best Local Similarity 25.5%; Pred. No. 32;
Matches 48; Conservative 28; Mismatches 83; Indels 29; Gaps 9;

Qy 5 NDTLVTADVRNGIDGH---ALADRLGLDEAETAWELSTGIDDDTMAALAAEQPLFATA 61
Db 1012 NVTLITAP--KGSDETFKAGTQQIGF--SNITPEIRTTDTATQWYLTGYQSVADARA 1067
Qy 62 DALVTDFYDHLSEYRTQDLFANSTKTVEQLKETQ-----AEVLLGLGRGE--YDTEYA 113
Db 1068 SKIATDFMD--SGYKSFTEVNNLKRMGDLRDSQDAGGAWRMNGTSGESGYRDNVT 1125
Qy 114 AQRARIGKIHVLGLGPDVYLGAATRY-----YTGLDLADADDVADRGEAAAV 164
Db 1126 HVQIGADRKHELNGI--DLFTGALLTYDNNASSQAFSGKTKSLGGVYAGLFGSGAYF 1183
Qy 165 DELVARFL 172
Db 1184 D-LIGKYL 1190

RESULT 10
Q81JUN0 PRELIMINARY; PRT; 433 AA.
AC Q81JUN0;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Methyl-accepting chemotaxis protein.
GN BA5673.
OS Bacillus anthracis (strain Ames).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

```

```

OX NCBI_TaxID=198094;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608414; PubMed=12721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.J., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA Thomson B., Friedlander A.M., Koshler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria.";
RL Nature 423:81-86(2003).
DR EMBL; AB017041; AAP29306.1; -.
DR TIGR; BA5673; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; P:signal transducer activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR004089; Chmtaxis trans.
DR InterPro; IPR004090; Me chemotaxis.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS01111; CHEMOTAXIS_TRANSDUC_2; 1.
KW Complete proteome.
SQ SEQUENCE 433 AA; 49391 MW; E06512BA0E696D92 CRC64;

Query Match 10.2%; Score 95; DB 16; Length 433;
Best Local Similarity 19.3%; Pred. No. 6.9;
Matches 32; Conservative 29; Mismatches 71; Indels 34; Gaps 3;

Qy 31 AEIARWLSFTGIDDDTMAALAAEQPLFATAVDYDHLSEYRTQDLFANSTK--- 87
Db 33 SELKVQMDMLHISKEDLQIVKVLQFFIYEIDWITEKFY-----ANITKPN 79
Qy 88 -----TVEQLKETQAEYLLGLGRGEYDTEYAAQARIGKIHVDVLGLGPDVYLGAAT 138
Db 80 LITTIERYSSIPKQLKTHIKELFGDMHEDFIEQRVKIARHVVQIGLHKRWYTAAYQ 139
Qy 139 RYTTGLDLADADDVADRGEAAAVDELVARFLPMLKLTFFDQ01 184
Db 140 ELFRSINKILTKI-----TTIDFSYSINVINKLFTLEQEL 176

RESULT 11
Q8W6U7 PRELIMINARY; PRT; 419 AA.
ID Q8W6U7;
AC Q8W6U7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Gp4.
GN 4.
OS Bacteriophage phiE125.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Lambda-like viruses.
OX NCBI_TaxID=180504;
RN [1]
RP SEQUENCE FROM N.A.
RA Woods D.E., Jeddleloh J.A., Fritz D.F., DeShazer D.;
RT "Burkholderia thailandensis B125 Harbors a Temperate Bacteriophage
RT Specific for Burkholderia mallei.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF447491; AAL40277.1; -.
DR InterPro; IPR006944; Phage portal.
DR InterPro; IPR006427; Portal HK97.
DR Pfam; PF04860; Phage_portal_1.

```

TGFRAMS; TIGR01537; portal_HK97; 1.
 SQ SEQUENCE 419 AA; 46404 MW; C15C3CC08F21D2C1 CRC64;
 Query Match 10.1%; Score 94.5; DB 9; Length 419;
 Best Local Similarity 21.7%; Pred. No. 7.2;
 Matches 49; Conservative 28; Mismatches 74; Indels 75; Gaps 10;
 Qy 5 NDTLVADVNGIDGHALADRIGLDRAETAWRLSFTGIDDDTWAALAAE-----QPLFEAT 60
 Db 201 NGTALSGVIERPKDAPALKDQASVDRTDGNNAKFGQGNKAKVALLQEGMTFRPLSMTN 260
 Qy 61 ADALVTD-----FYDHL-----SYE 76
 Db 261 VDAALIDALRLSALDIARIYKIPAHMWNELETRATFSNIEHQSLQFVIYTLPLPWKRHEQA 320
 Qy 77 RTQDLFANSTKTVQKQAEY-LIGLGRGEYDTEYAAQARIGK-----IHDVLGI- 128
 Db 321 KTRDLLPSER-----KQYFIEYNLGALLRGDQSSRYAA--AVGRQWGLSINDIRRL 373
 Qy 129 -----GPDVYLGAITYTYTGLLDA-LADDVADRGEEAAAADDEL 167
 Db 374 NMPPVKGDIYLSF-----NMNVDAKPKQOLPVKSEPTKKAALDEI 414
 RESULT 12
 Q9HRW3 PRELIMINARY; PRT; 883 AA.
 ID Q9HRW3
 AC Q9HRW3, 2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Vng0514C.
 GN Vng0514C.
 OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
 OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 EN NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20504483; PubMed=11016950;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonksi P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Beck R.F., Pohlshocher M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
 RA "Genome sequence of Halobacterium species NRC-1.";
 RT Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 DR EMBL; AB005003; AAG19045.1; -.
 DR PIR; A84210; A84210.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
 DR GO; GO:0007059; P:chromosome segregation; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003439; ABC transporter.
 DR InterPro; IPR003405; SMC_C.
 DR InterPro; IPR003395; SMC_N.
 DR InterPro; IPR000533; Tropomyosin.
 DR Pfam; PF02483; SMC_C; 1.
 DR Pfam; PF02463; SMC_N; 1.
 DR PRINTS; PR00194; TROPOMYOSIN.
 KW Complete proteome.
 SQ SEQUENCE 883 AA; 96969 MW; F8284D3326C92BB9 CRC64;
 Query Match 10.1%; Score 94.5; DB 17; Length 883;
 Best Local Similarity 28.3%; Pred. No. 21;
 Matches 51; Conservative 23; Mismatches 63; Indels 43; Gaps 10;
 Qy 12 DVNCGIDGH--ALADRIGLDRAETAWRLSFTGIDDDTWAALAAEQPLFEA-----TAD 62

B
1011K